

基于相关滤波的目标跟踪算法研究综述

孟晓燕, 段建民

(北京工业大学信息学部, 北京 100124)

摘要: 基于相关滤波理论的判别式跟踪方法由于其高效性和鲁棒性, 已经取得了一系列的进展, 成为了目标跟踪领域的研究热门. 为了使更多国内外学者对相关滤波目标跟踪理论及其发展进行进一步研究与探索, 对该领域研究现状进行综述. 首先, 介绍了相关滤波理论及其用于实现目标跟踪任务时的一般框架, 并重点描述了典型的核相关滤波跟踪方法. 其次, 讨论了目标跟踪技术应用于实际场景时面临的诸多难题, 详细分析了特征表示和自适应尺度更新这2个主要难点. 然后, 从基本类相关滤波、部件类相关滤波、正则化类相关滤波和 Siamese 网络类相关滤波这4个类别对具有代表性的算法进行分析与讨论, 并指出了未来可能的发展趋势. 最后, 在 OTB2013 和 OTB100 基准数据集上对 32 种相关滤波类跟踪算法就精确度、成功率和帧率进行了对比, 在 VOT2017 数据集上对 10 种相关滤波类跟踪算法就平均重叠期望 (expected average overlap, EAO)、Accuracy 和 Robustness 三个性能指标进行了对比, 体现了相关滤波跟踪器 (correlation filter trackers, CFTs) 的优越性. 尽管相关滤波理论在目标跟踪领域具有广阔的应用前景, 但是复杂场景和自身因素的影响导致其仍然是一个极具挑战性的研究方向, 研究兼备准确性与鲁棒性的 CFTs 对于目标跟踪领域的发展具有重要意义.

关键词: 机器视觉; 目标跟踪; 相关滤波; 特征表示; 尺度更新; 孪生网络; 性能指标

中图分类号: TP 391.4

文献标志码: A

文章编号: 0254-0037(2020)12-1393-24

doi: 10.11936/bjtxb2019030011

Advances in Correlation Filter-based Object Tracking Algorithms: A Review

MENG Xiaoyan, DUAN Jianmin

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China)

Abstract: In the field of object tracking, the discriminant method based on correlation filter theory has made a series of advances and becomes a hot research topic because of its efficiency and robustness. The current research statuses of the tracking field were reviewed to allow more scholars to explore the theory and development of correlation filter-based trackers. First, the correlation filter theory and the general framework for object tracking were introduced, and the classical kernelized correlation filter was described in detail. Second, the difficulties of object tracking technology when applied in the real application were discussed, and the main difficulties of feature representation and adaptive scale updating were analyzed in detail. Then, the representative algorithms were analyzed and discussed from the four categories of basic correlation filter, part correlation filter, regularized correlation filter, and Siamese network correlation filter, and the possible future development trend was pointed out. Finally, 32 types of correlation filter-based trackers were compared in terms of accuracy, success rate and frame rate on the OTB2013 and

收稿日期: 2019-03-18

基金项目: 北京市自然科学基金资助项目 (JJ002790200802); 北京市属高等学校人才强教计划资助项目 (038000543117004)

作者简介: 孟晓燕 (1990—), 女, 博士研究生, 主要从事智能交通、图像处理、机器视觉方面的研究, E-mail: 529108419@

qq.com

OTB100 standard data sets, and 10 types of correlation filter-based trackers were compared in terms of EAO, and the accuracy and robustness on the VOT2017 data set, further indicated the advantages of correlation filter tracking algorithms. The research on correlation filter theory has extensive applications in the object tracking field. However, it is still a challenging research direction due to the influence of complex scenes and their own factors. Developing a highly efficient and robust correlation filter tracker is considered significant.

Key words: computer vision; object tracking; correlation filter; feature representation; scale updating; siamese network; performance indicator

目标跟踪是一个机遇与挑战并存的机器视觉研究方向,因在自动驾驶^[1]、交通流监控^[2]、人机交互^[3]、机器人^[4]、医学诊断^[5]和行为识别^[6]等领域得到广泛应用,而成为了研究热点. 机遇主要体现在虽然经过十几年的研究,该方向已经有了许多突破性的进展,但是仍然存在进一步深入研究和提升的空间. 挑战性则体现在复杂环境和目标本身 2 个方面,具体而言,跟踪失败可以归因于跟踪过程中发生的遮挡、光照变化、尺度和形态差异以及快速运动等不利因素.

目标跟踪旨在从连续视频帧中识别出感兴趣区域. 一般而言,跟踪框架由 4 个模块组成,分别是目标初始化、外观建模、运动估计和目标定位. 目标初始化是对目标的位置或感兴趣区域标示包围框的过程,外观建模是由具备判别能力的目标特征和构造的数学模型组成的检测过程,运动估计是推断跟踪后续帧中目标所处位置的过程,而目标定位操作则涉及到最大后验概率预测或贪婪搜索策略. 大量的跟踪算法被相继提出,回答了何为视觉目标跟踪,何种学习机制适用于跟踪任务,以及如何有效定义外观模型和运动模型等问题. 文献[7-11]从不同的侧

重点对跟踪算法进行了调查,但是考虑到近年来该领域的迅速发展,以及相关滤波和深度学习在完成跟踪任务时的突出表现,目前仍缺乏一个针对性较强且相对全面的论述. 本文旨在提供一个基于相关滤波理论的目标跟踪方法的研究进展综述,希望可以为读者们就跟踪算法的多样性提供一个有组织、有层次的参考文献,并为其未来将要开展的研究工作提供有价值的参考思路.

本文工作如下:首先对相关滤波跟踪算法进行了介绍,并且对核相关滤波器进行了详细描述. 然后就视觉目标跟踪面临的主要难题、目前的研究现状等方面展开分析与讨论. 最后在 OTB2013、OTB100 和 VOT2017 基准数据集上,对具有代表性的相关滤波类跟踪算法进行了对比,进一步说明了其优越性.

1 相关滤波类跟踪算法

基于相关滤波理论的跟踪方案在频域内进行计算有效控制了运算成本,提高了跟踪效率. 依据已有知识,可将相关滤波跟踪器 (correlation filter trackers, CFTs) 的一般框架归纳为图 1. 首先在初始

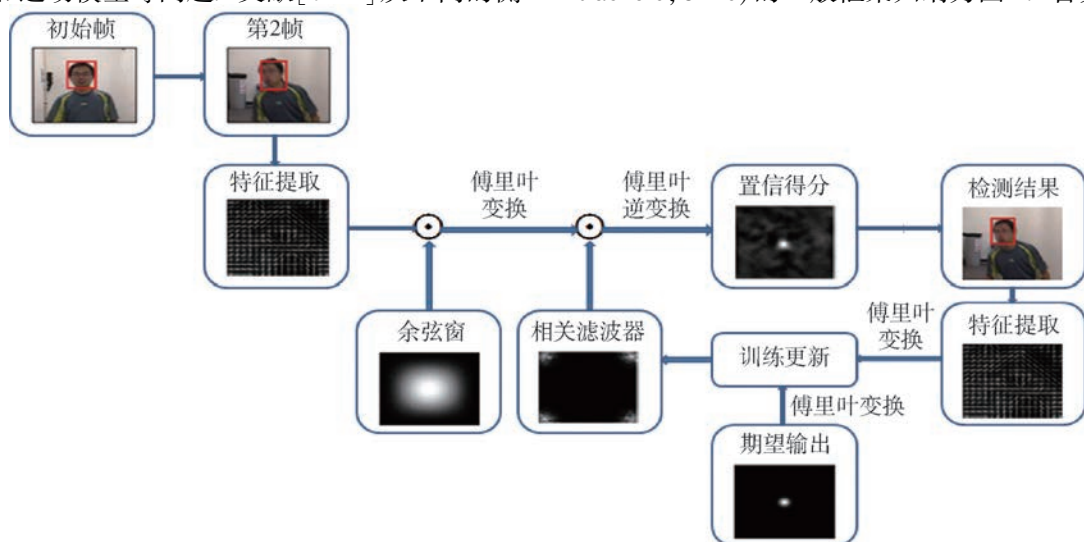


图 1 相关滤波跟踪算法框架示意图

Fig. 1 Illustration of the framework of correlation filter-based tracking algorithms

帧确定的目标位置提取图像块,训练滤波器.然后在跟踪过程中,根据前一帧中的目标位置,在当前帧中估计包含目标的图像块.为了有效地表示目标外观,可在选定的图像块中提取手工特征、深度特征或融合特征,并利用余弦窗口平滑边界.通过离散傅里叶变换执行相关滤波操作.最后通过傅里叶逆变换得到响应地图,响应得分最大值所在的位置即为目标在当前帧中的新位置.

1.1 相关滤波理论

设 \mathbf{h} 为相关滤波器, \mathbf{x} 为输入数据,可以取图像特征或原始像素.根据卷积定理,空域中的卷积等于频域中元素间的乘积,可得

$$\mathbf{x} \otimes \mathbf{h} = F^{-1}(\hat{\mathbf{x}} \hat{\mathbf{h}}^*) \quad (1)$$

式中: \otimes 为卷积; \odot 为元素间的乘法运算; F^{-1} 为傅里叶逆变换; $*$ 为复共轭.式(1)给出了输入 \mathbf{x} 和滤波器 \mathbf{h} 之间的置信地图,选择目标周围置信度值最大处对相关滤波器进行更新.假设 \mathbf{y} 为期望输出,对于新目标外观 \mathbf{z} ,相关滤波器 \mathbf{h} 需满足

$$\mathbf{y} = F^{-1}(\hat{\mathbf{z}} \hat{\mathbf{h}}^*) \quad (2)$$

因此,可得

$$\hat{\mathbf{h}}^* = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{z}}} \quad (3)$$

式中 $\hat{\mathbf{y}}$ 为频域内的期望输出.快速傅里叶变化(fast Fourier transformation, FFT)有效降低了计算成本,对于 $n \times n$ 大小的图片,循环卷积运算的复杂度为 $O(n^4)$,而 FFT 只需 $O(n^2 \lg n)$.

1.2 核相关滤波器

文献[12]将 CFTs 表示为求解岭回归问题,利用循环矩阵对样本进行密集采样,并引入核函数提升分类器的性能,具体的实现过程可以分为训练、检测和更新3个部分.

1) 训练.在第 t 帧中,以检测到的目标位置为中心选取大小为 $M \times N$ 的图像块 \mathbf{x} 作为基样本,采用循环移位得到训练样本 \mathbf{x}_i ,其中 $i \in \{0, 1, \dots, M-1\} \times \{0, 1, \dots, N-1\}$.对于递归最小二乘法(recursive least square, RLS)分类器,分类实质为找到 \mathbf{w} 使得函数 $f(\mathbf{x}) = \mathbf{w}^T \mathbf{x}$ 在 \mathbf{x}_i 上的响应与 \mathbf{y}_i 的平方误差值最小,即

$$\min_{\mathbf{w}} \sum_i (f(\mathbf{x}_i) - \mathbf{y}_i)^2 + \lambda \|\mathbf{w}\|^2 \quad (4)$$

式中: \mathbf{w} 为分类器参数; $\lambda \|\mathbf{w}\|^2$ 为正则化项,其目的是为了防止滤波器过拟合.通过非线性映射函数将 \mathbf{w} 表示为 \mathbf{x}_i 的高维特征矩阵 $\phi(\mathbf{x}_i)$ 的线性加权形式,即 $\mathbf{w} = \sum_i \alpha_i \phi(\mathbf{x}_i)$,式中 α 为 \mathbf{w} 的对偶空间变

量.利用核函数 $\kappa(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$ 解决高维特征矩阵点积计算问题,分类器可以进一步表示为 $f(\mathbf{x}) = \sum_i \alpha_i \kappa(\mathbf{x}_i, \mathbf{x})$,此时 \mathbf{w} 的求解转换为 α 的求解问题.因为核函数是循环矩阵,则通过

$$\hat{\alpha} = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{k}}^{xx} + \lambda} \quad (5)$$

计算 $\hat{\alpha}$.式中核相关 $\hat{\mathbf{k}}^{xx}$ 为 $\kappa(\mathbf{x}_i, \mathbf{x})$ 的第 i 个元素,文献[12]中使用的核函数为高斯核,计算公式为

$$\mathbf{k}^{xx'} = \exp \left(-\frac{1}{\sigma^2} (\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2F^{-1}(\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}}')) \right) \quad (6)$$

式中: σ 为核函数参数; $\hat{\mathbf{x}}^*$ 为 $\hat{\mathbf{x}}$ 的复共轭.

2) 检测.样本训练完成之后,对于新读入的视频帧,选取以上一帧中估计的目标位置为中心、大小为 $M \times N$ 的图像块上进行检测,得到的响应图表示为

$$f(\mathbf{z}) = F^{-1}((\hat{\mathbf{x}}^z) \odot \hat{\alpha}) \quad (7)$$

$f(\mathbf{z})$ 得分最大值处所对应的位置即为当前帧中目标的位置.

3) 更新.为了适应跟踪过程中目标外观的变化,采用线性插值的方式对滤波器和目标外观模型进行更新,更新方式为

$$\begin{cases} \hat{\alpha}^t = (1 - \theta) \hat{\alpha}^{t-1} + \theta \hat{\alpha} \\ \hat{\mathbf{x}}^t = (1 - \theta) \hat{\mathbf{x}}^{t-1} + \theta \hat{\mathbf{x}} \end{cases} \quad (8)$$

式中: t 为帧数; θ 为学习率.

2 视觉目标跟踪的难点

由于实际场景的复杂性以及目标自身的变化,视觉目标跟踪面临很多技术难点,本节主要从特征表示与尺度变化2个方面进行分析.与其他视觉任务一样,对于目标跟踪而言,特征表示直接关系到目标外观模型的有效性,进而影响跟踪的成败.因此,选择合适的独立特征或融合特征是设计跟踪框架时需要重点考虑的问题.此外,目标在视频序列中的尺度会随着目标与摄像机之间的距离发生改变,变大或变小均会对之后的建模产生不利影响,累计误差导致跟踪失败.

2.1 特征表示

最初将相关滤波理论引入目标跟踪领域的 MOSSE^[13]跟踪器和基于检测的核循环结构 CSK^[14]跟踪器采用的均为单通道灰度特征,而灰度特征的代表能力不足以处理背景复杂或者目标与背景颜色相似的情况.之后 Henriques 等^[12]将单通道特征扩

展为多通道方向梯度直方图特征,利用图像的梯度信息提升算法的跟踪性能.文献[15]将 Gray 扩展为多通道颜色名称(color names, CN)特征,结合自适应降维策略,在降低计算成本的同时提升了跟踪性能.方向梯度直方图(histogram of oriented gradients, HOG)特征和颜色特征在描述目标方面显示出了明显的优势,所以之后提出的许多算法采用的均为 HOG 特征或 CN 特征,如 RPT^[16]、DSST^[17]、FDSST^[18]和 LCMF^[19]等.

考虑到利用单一特征描述目标的局限性,研究者们将关注点放在了如何融合不同特征的特性实现优势互补上,但是复杂特征不可避免地增加了计算复杂度,因此如何在保证跟踪实时性的同时提升跟踪性能是需要重点考虑的问题.文献[20]从优化特征表示的角度出发提出了 SAMF 跟踪器,同时对 Gray、CN 和 HOG 特征进行建模,提升了复杂环境下跟踪器的鲁棒性.文献[21]从特征间差异性角度出发,利用深度信息和颜色信息

的优势互补,提出了 DS-KCF 跟踪器,保证实时性的同时,有效提升了算法性能.受 HOG 强大表征力的启发,文献[22]将一种新的描述算子 MC-HOG 与相关滤波框架结合提出了 MOCA 跟踪器,利用多颜色通道的方向梯度直方图特征进行建模,在一定程度上提升了跟踪器的性能.文献[23]提出的 STAPLE 算法通过保持 2 个独立的回归问题来开发每个补丁的固有结构,采用 HOG 和全局颜色直方图共同描述目标.在颜色模板中,前景和背景区域由先前的估计位置计算,得到搜索区域中每个像素的得分,使用积分图计算颜色响应地图.在 HOG 模板中,在先前估计位置确定的搜索区域上提取 HOG 特征,并将其与 CF 卷积得到密集响应模板,最终通过 2 个模板得分的线性组合估计出目标的位置.它可以视为 DSST 和 DAT 的互补与结合,2 种算法高效无缝结合之后,不仅准确性得以提升,其计算效率也有了显著的提升,帧率达到了 80 帧/s,STAPLE 的算法框架如图 2 所示.

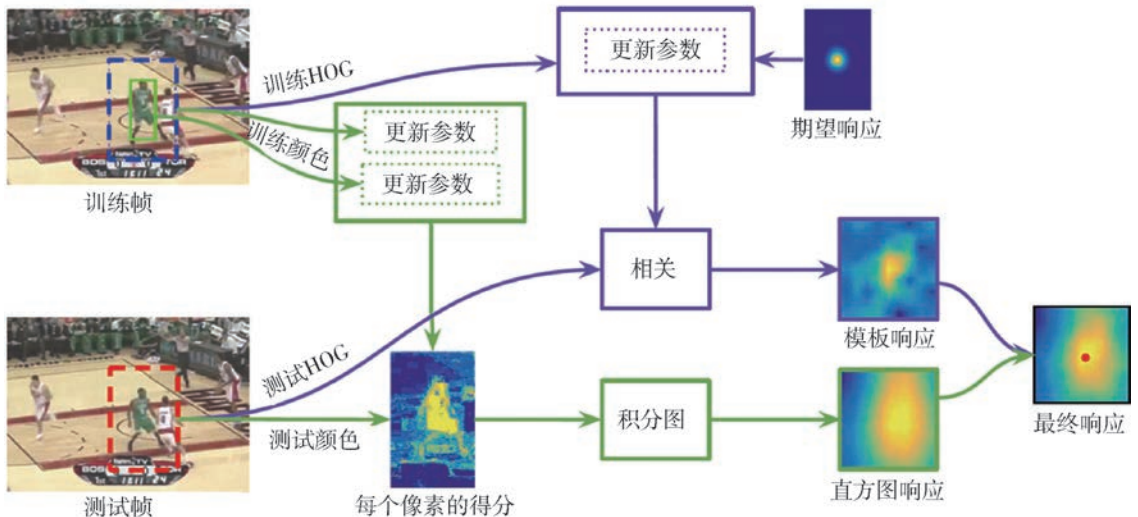


图2 STAPLE 的算法框架

Fig. 2 Algorithm framework of STAPLE

随着深度学习的快速发展,许多结合深度卷积特征的 CFTs 应运而生.文献[24]在相关滤波框架中引入丰富的层次卷积特征用于视觉跟踪,基于 KCF 框架,采用在 ImageNet 上使用 VGG-19 训练好的特定三层特征替代原始 HOG 特征进行建模,有效提升了跟踪性能.文献[25]在 SRDCF 框架的基础上结合卷积特征进行建模提出了 DeepSRDCF 跟踪算法.不同卷积层的特征适于完成不同的任务,低层特征利于定位,而高层特征因包含丰富的语义信息适于分类. CFNet^[26]对比了所提算法框架分别使

用 conv1、conv2 和 conv5 时的性能差异,得出了 CFNet-conv2 更适于完成跟踪任务的结论.之后提出的许多方法以不同的方式结合深度特征构造跟踪框架,获得了优异的成绩,但是此类滤波器忽略了卷积层之间的关系,所以仍然存在一定的局限性. Danelljan 等^[27]提出的 C-COT 使用神经网络 VGG-Net 提取深度特征,将原始的彩色图像和 2 个卷积层的输出作为特征,虽然提升了准确度,但是复杂的特征严重影响了计算速度,难以满足实时性的要求. ECO^[28]在其基础上通过因式分解操作,将 HOG、CN

和卷积神经网络(convolution neural network, CNN)的维度进行了不同程度的降维,减少了训练参数,进而有效降低了计算复杂度. ECO + [29]将特征网络换为 ResNet-50,采用数据增强扩充训练集,将浅层特征和深度特征分而治之,最终两部分响应自适应融合得到最优结果. 作者具体分析了浅层特征和深度特征的特征差异,得出了深度特征偏向于鲁棒性而手工特征更关注准确度的结论.

由上述讨论可知,CFTs经历了从单通道特征到多通道特征、从单一特征到融合特征、从人工特征到深度特征的过程,如何为跟踪任务匹配合适的特征以及采取何种方式进行特征融合是实现跟踪器性能提升急需解决的关键问题.

2.2 自适应尺度更新问题

传统 CFTs 通常采用固定大小的窗口,以致于在跟踪过程中易发生目标尺度增大、边缘信息丢失或尺度减少引入背景干扰等问题. 为了更好地处理目标尺度变化带来的跟踪漂移问题,许多学者从不同的角度进行了研究. Li 等[20]针对尺度问题提出了 SAMF 方法,该方法在 KCF 跟踪框架的基础上,将滤波器在多尺度缩放的图像块上进行检测得到响应地图,取响应最大的位置为目标位置与最佳尺度. Danelljan 等[17]提出的 DSST 算法同样采取尺度穷举的思路,与 SAMF 不同的是其采用的是分步穷举策略,即先进行平移跟踪,后进行尺度跟踪. SAMF 可以看作是一个全局寻优的过程,而 DSST 可以看作是局部寻优的过程,全局最优并不能保证局部最优,所以 SAMF 的得分最高处并不能保证尺度最优. 此外,与 SAMF 统一选取 HOG + CN 特征建模不同,DSST 的分步策略为不同滤波器的特征描述子提供了更加灵活的选择. 独立尺度滤波器的引入,不可避免地增大了计算复杂度,而且相较于 SAMF 的 7 个粗略尺度,DSST 的尺度池因包含了 33 个精细尺度,降低了计算效率. 之后提出的加速版 FDSST[18]就是考虑到复杂度的问题,采用降维操作和 QR 分解降低了计算量. 文献[30]提出的多模板尺度自适应核 CFT-CF + MT,与文献[17]和[20]采用的策略类似,也是在一定的尺度池中进行穷举搜索,使用后验概率取代最大似然概率,获得了健壮的跟踪性能. 文献[31]借鉴配置独立尺度滤波器的思路,将分步策略化为同步进行,得到目标最佳位置以及尺度. 考虑到连续帧中跟踪目标的尺度变化范围较小,采用上述穷举尺度池的方法一定程度上满足了尺

度更新的需要,而且配置独立尺度滤波器的方法,具备很强的移植能力,为解决尺度变化问题提供了一种可行的思路. 考虑到尺度滤波器的使用对于耗时的负面影响,Solis 等[32]基于 KCF 框架提出了一种快速可伸缩方案,通过引入可调高斯核函数和帧间关键点匹配技术处理尺度问题,并从特征描述子、模板大小等方面综合降低了运行时间,是一种高效的 CFT.

除了穷举尺度池的方法,分块处理也是实现尺度自适应的一种有效方式. 文献[16]提出的可信块跟踪器 RPT,由部件间的相对位置关系推断目标的尺度变换情况. 文献[33]提出了一种实时分块自适应核相关滤波器 RPAC,将跟踪目标分解为 5 个部件,为每个部件配备独立的滤波器,通过计算各个响应地图中响应得分最大处的变化来估计目标尺度的变化. 文献[34]从平衡准确性与实时性的角度出发提出了 DPCF 跟踪器,通过一个全局滤波器和多个部件滤波器间的耦合作用协同处理局部遮挡和尺度变化问题.

除了早期的跟踪器,之后提出的跟踪框架大多包含了尺度适应策略. 其中,尺度穷举法和分块模型是实现尺度自适应的 2 个主要思路,如何在尽量不影响执行速度的情况下,确定尺度因子及空间结构是需要重点考虑的问题.

3 研究现状分析

目标跟踪之所以具有极大的挑战性,是因为以下几方面因素[35]:1) 目标外观信息的缺失;2) 背景噪声的干扰;3) 不可避免的外观变化. 尽管目前的研究已经在建立外观模型和鲁棒跟踪方面取得了明显进展,目标跟踪技术应用于实际场景时仍然是一个非常复杂的问题. 除了第 2 节中讨论的特征表示以及尺度自适应更新方面,像目标遮挡、跟踪漂移等问题,也在很大程度上影响了跟踪性能.

本节针对上述挑战和问题,分别从基本类 CFTs、部件类 CFTs、正则化类 CFTs 和 Siamese 网络类 CFTs 共 4 个类别对已有的目标跟踪算法进行分析与讨论.

3.1 基本类 CFTs

基本类 CFTs 以 KCF 为基本框架,旨在改进 KCF 自身的缺陷以及处理不同的跟踪难题. 具体而言,这些跟踪器从特征表示、尺度变化的处理、核函数、长期跟踪、响应分布以及算法集成等角度进行优化,以实现跟踪性能的提升.

从特征优化的角度出发,文献[24]在相关滤波框架中引入了丰富的层次卷积特征用于视觉跟踪,提出了 HCFT 跟踪器. HCFT 在 KCF 框架中引入在 ImageNet 上使用 VGG-19 训练好的特定三层特征替代原始 HOG 特征,并为每一层特征配置独立的相关滤波器进行模板学习,得到置信图后进行加权融合获取目标位置. 目标定位时,兼顾了低层特征的高分辨率优势和高层特征的语义信息,提升了跟踪的精确度和鲁棒性. Ma 等^[36]对 HCFT 进行了扩展,加入重检测模块和尺度估计,提出了基于层次关联特

征的跟踪器 HCFT+. 文献[37]学习卷积层输出的 CF 编码目标外观,推断每一层的响应,并以由粗到细的方式进行目标定位. 算法增加额外的 CF 用以保持目标外观的长期记忆,而且将其应用于不同类型的目标方案. Qi 等^[38]提出的 HDT 算法同样在特征层面进行了研究,集成了 CNN 不同层特征的优势,首先通过相关滤波器在卷积层建立弱跟踪器,并为每一个弱分类器分配权重作为决策置信,然后采用 boosting 的思想进行对冲生成强跟踪器,得到最终的跟踪结果,HDT 跟踪框架如图 3 所示.

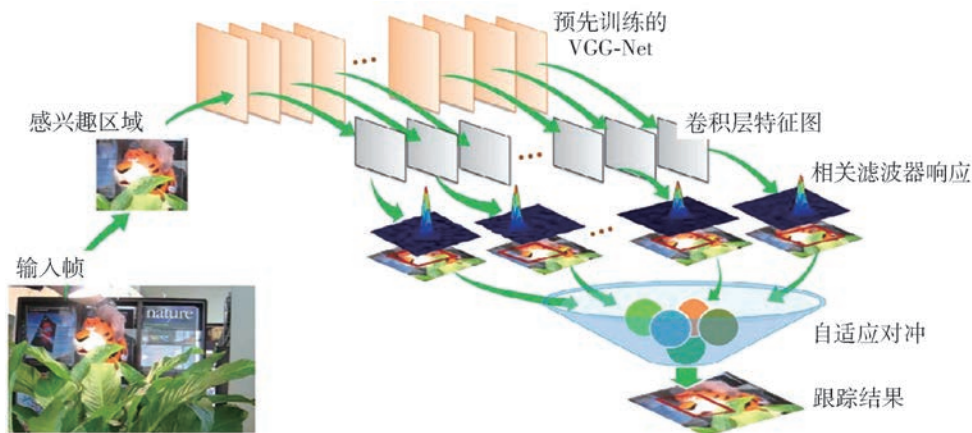


图3 HDT的算法框架

Fig. 3 Algorithm framework of HDT

从处理尺度变化的角度出发,Zhang 等^[39]在 KCF 基本框架上提出的 RAJSSC 算法更能够同时从空间位移、尺度变化和旋转变换对目标外观变化建立模型. 算法首先通过循环矩阵在尺度和位置空间执行穷举搜索,然后将目标模板从笛卡儿坐标系转移到对数极坐标系,确保目标发生旋转时,仍能保持其循环结构. Li 等^[20]针对尺度问题提出了 SAMF 方法,将滤波器在多尺度缩放的图像块上进行检测得到响应地图,取响应最大的位置为目标位置与最佳尺度. Danelljan 等提出的 DSST^[17]算法和加速版 FDSST^[18]同样采取尺度穷举的思路对尺度变换进行了处理,具体的实现方式已在 2.2 节中给出. 类似文献[17-18,20]采用的穷举尺度池方法一定程度上满足了尺度更新的需要,为解决尺度变化问题提供了一种可行的思路. 考虑到尺度滤波器的使用对于耗时的负面影响,文献[32]基于 KCF 框架提出了一种快速可伸缩方案,通过引入可调高斯核函数和帧间关键点匹配技术处理尺度问题. 文献[40]将 Edgeboxes 引入 KCF 框架提出了 KCFDPT 方法,通过生成目标候选区域应对目标尺度和长宽比的变化.

从核函数的角度出发,文献[41]针对基本相关滤波器只使用单核的问题,提出一种基于多核相关滤波器的跟踪算法,算法充分利用不同特征的幂频谱(power spectrums)及其判别不变性来提升性能,并且采用最优二分搜索和特征快速估计进行尺度估计,同时利用特征金字塔的最小层数,有效降低了计算量.

从实现长期跟踪的角度出发,文献[42]在 DSST 的基础上,引入了负责目标置信度的相关滤波器,即通过平移 CF、尺度 CF 和置信度 CF 共同实现目标的跟踪任务. 如图 4 所示,LCT 通过对时间上下文相关信息建模实现平移估计,采用外观信息构建尺度金字塔实现尺度估计,使用在线随机 ferns 检测器实现目标丢失情况下的重新检测,很大程度上提升了目标大面积遮挡和移除视野情况下跟踪的鲁棒性. 作者在 LCT 基础上,使用支持向量机(support vector machine,SVM)替代在线随机 ferns 检测作为重检测模块提出了 ILCT^[43]方法. 文献[44]受认知心理学 ASMM(Atkinson-Shiffrin memory model)启发提出了多存储跟踪器 MUSTer,将短期存储和长期存储聚合到图像信息并进行跟踪. 短期存储涉及到一个集成

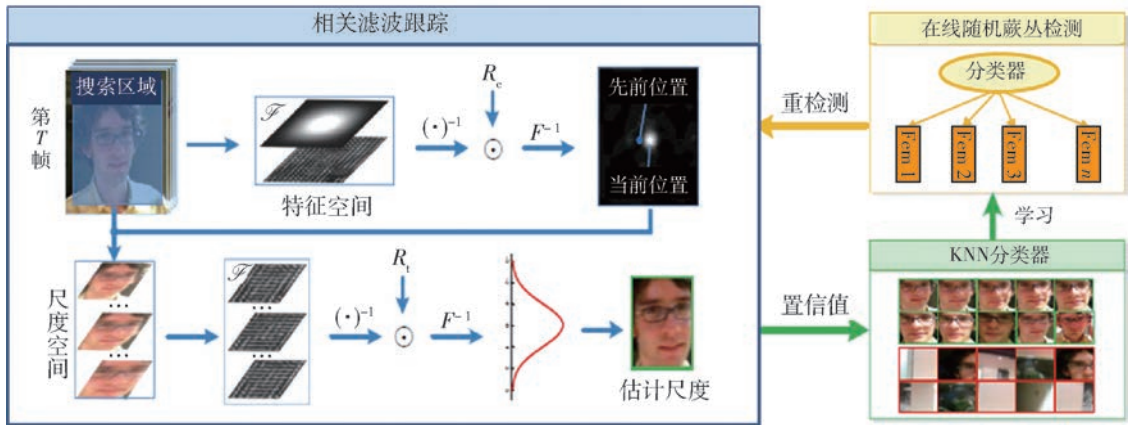


图 4 LCT 的算法框架

Fig. 4 Algorithm framework of LCT

相关滤波器 ICF 用于结合时空一致性,而长期存储涉及到随机抽样一致算法 (random sample consensus, RANSAC) 估计和关键点匹配跟踪控制输出。

从响应分布的角度出发,文献[45]为克服单中心高斯分布作为目标响应会降低跟踪性能的缺陷,提出了一种可以自适应更改目标响应的通用框架,以降低跟踪器循环位移不能可靠近似转换时的敏感性。文献[46]基于滤波器响应的各向异性,提出了3种稀疏损失函数用以减少滤波器的损耗及过拟合情况。文献[47]同样针对 KCF 滤波器的响应分布问题,提出了一种输出约束转移方法,在贝叶斯优化框架中对相关响应的分布进行建模,缓解了跟踪的漂移问题。一方面,OCT 基于对目标的相关响应服从高斯分布的合理解释,利用高斯分布选择训练样本,降低了模型的不确定性。另一方面,OCT 基于将数据分布转换为优化变量约束的理论,提升了估计效率。

从算法集成的角度出发,由于每种跟踪算法之间都存在一定的差异,它们根据被跟踪目标以及环境的不同进行假设,因此不同算法均能体现出自身的优势和不足,对不同算法进行集成实现优势互补是提升跟踪性能的一种有效方式。文献[48]提出了一种基于支持相关滤波器的跟踪器 SCF。算法通过循环矩阵对 SVM 模型进行重构,将交替优化过程与离散傅里叶变化结合,即将跟踪问题转换为支持相关滤波器的迭代学习过程,进而实时找到全局最优解。文献[19]提出的 LCMF 算法,同样使用结构 SVM 作为分类器,引入相关滤波器的速度优势,通过多峰检测避免了相似物体和背景的干扰,并且采用高效的模型更新策略,兼顾准确性的同时大大加快了计算效率。文献[49]提出了一种多任务相关粒子滤波器进行视觉跟踪,所提的 MCPF 跟踪框架结

合了 MCF 和粒子滤波器的优势,利用不同特征之间的相关性联合学习相关滤波器,通过粒子抽样策略有效处理了尺度变化问题,并引导抽样粒子向目标状态分布的模式方向运动以提升跟踪性能。文献[50]提出的并行跟踪与验证跟踪框架从寻求跟踪性能和效率间平衡点的角度出发,将跟踪过程分为跟踪器和验证器2个部分。跟踪部分采用 FDSST 算法,负责计算实时推理和估计跟踪结果,验证器部分负责多尺度验证结果,其中跟踪部分和验证部分采用2个线程工作异步进行。

3.2 正则化类 CFTs

判别式相关滤波器的跟踪性能主要受到3个方面的限制:1) 滤波器尺寸需要和块尺寸相等的条件限制了检测范围。2) 固定的搜索区域造成了训练样本集中负样本的缺失,产生模型过拟合的情况,当跟踪发生遮挡时难以实现重新检测。3) 采用周期假设产生边界效应问题。虽然选择较大的搜索区域可以解决此类问题,但是引入过多的背景信息又会降低跟踪器的判别能力,因此,研究者们引入正则化策略来改善这一问题,许多正则化类相关滤波器应运而生。

针对判别相关滤波器采用周期假设导致不必要的边界效应这一问题,文献[51]在 DCF 框架中引入正则项提出了空间正则化判别 CFT-SRDCF。算法根据空间位置确定滤波器的惩罚系数,弱化了背景信息的干扰,并利用高斯-赛德尔迭代方法进行在线训练,提升了跟踪性能。SRDCF 对尺度变化的处理方法采用了 SAMF^[20] 的尺度金字塔法。通过引入空间权重函数对目标边界框外的非零滤波器进行惩罚的思路是合理的,但是权重在整个序列中是固定的,只能抑制背景,却不能随着形状的变化而增强目标。文献

[52]在判别相关滤波框架中结合了颜色概率,提出了基于通道可靠性和空间置信度的 CSR-DCF 跟踪方法. 一方面,空间置信度地图支持滤波器自适应选择适于跟踪的目标区域,减少了不必要的边界效应和矩形假设的局限性. 另一方面,CSR-DCF 框架对多通道特征直接求和的方式进行了优化,首先根据约束最小二乘解的性质估计得到通道可靠性,将其作为加权系数,然后采用自适应加权求和策略得到最终的响应结果,其算法框架如图 5 所示. SRDCF 和 CSR-DCF 都专注于惩罚滤波器,其中 SRDCF 是引入固定的权值来抑制边界处的滤波器,而 CSR-DCF 采用二值分割模型来消除对应背景的滤波器系数. 文献[53]从不同于 SRDCF 和 CSR-DCF 的思路出发,通过构造自适应特征权值来为像素分配不同的权值,兼备背景抑制和目标增强,提出了 FWDCF 跟踪器. FWDCF 的一个主要贡献是提出了伪相关滤波变量,将自适应权重整合于 DCF 框架,并采用高斯-赛德尔法在傅里叶域进行优化求解.

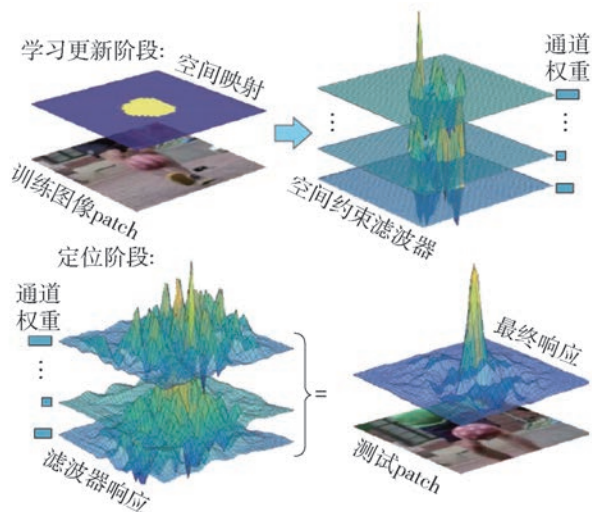


图5 CSR-DCF 的算法框架

Fig. 5 Algorithm framework of CSR-DCF

Danelljan 等^[54]将 SRDCF 框架在样本质量层面进行了优化发展为 SRDCFdecon, 通过一个联合外观模型和样本质量权重的统一公式实现了损坏样本的降低和正确样本的增加. 文献[25]使用 CNN 特征替代了原框架中采用的手工特征, 提出了 DeepSRDCF 算法, 并且证明了浅层卷积层与深层相比更适于完成视觉跟踪任务. 随着深度运动特征在动作识别领域的广泛应用, 文献[55]将深度运动特征和手工特征融合共同作用于 SRDCF 框架, 提出了 DMSRDCF 算法, 因深度运动特征为外观线索提供了互补信息而显著提升了跟踪性能.

文献[56]对 DMSRDCF 的工作进行了扩展, 具体分析了深度运动特征对视觉目标跟踪的影响及其适用条件.

在在线被动攻击学习的启发下, 从平衡主动学习和被动学习的角度出发, 文献[57]将时间正则化引入到 SRDCF 框架中, 提出了时空正则化相关滤波器 STRCF. 该框架不仅可以合理地逼近多幅训练样本上的 SRDCF 形式, 而且在发生巨大形变时体现出了更强的鲁棒性. STRCF 模型是凸的, 采用交替方向乘法 (alternating direction method of multiplier, ADMM) 对每个子问题进行求解时均能得到闭合解, 而且能在迭代次数较少时实现收敛, 得到最终的跟踪结果. 文献[58]在时空正则化相关滤波框架中引入双色聚类直方图模型提出了 CSCT 跟踪方法. 首先, 采用 K-means 算法对第一帧中目标真值的颜色通道进行聚类, 生成一个数据自适应非均匀量化器用于设计颜色直方图, 从而得到更加鲁棒的颜色模型. 然后, 为了避免外观突变造成的跟踪漂移, CSCT 框架提出了一种时空正则化方法来学习滤波器. 最后, 将颜色聚类直方图模型与时空正则化相关滤波模型进行线性组合, 得到最终的外观模型.

文献[27]提出的 C-COT 方法使用 VGG-Net 提取特征, 通过隐式插值将不同分辨率的特征图插值到连续空间域, 应用 Hessian 矩阵获得亚像素精度的目标位置. 针对 C-COT 使用高维特征训练大量滤波器而且每帧更新造成计算量增加的缺陷, 文献[28]在 C-COT 基础上, 提出了改进版本的 ECO 跟踪框架, 从 3 个方面对 C-COT 进行了优化. 首先, ECO 通过矩阵分解卷积操作对滤波器进行降维, 得到了一组具有显著能量的紧凑型滤波器. 其次, 从图 6 给出的 C-COT 和 ECO 的训练样本可看出, C-COT 的训练集由一系列连续的样本组成, 这样既产生了大量的冗余样本又可能导致对近期样本的过度拟合, 而 ECO 将样本划分为一定数量具备类内相似性和类间差异性的组件, 利用混合高斯模型对组件外观进行表示, 在降低样本数量的同时提升了样本的判别能力. 最后, ECO 采用稀疏型模型更新策略进一步降低了计算复杂度. 针对 ECO 不能受益于更好更深的深度特征这一反常现象, 文献[29]提出了 ECO+ 算法. 在 ECO 的基础上, ECO+ 将特征网络换为 ResNet-50, 采用数据增强扩充训练集, 将浅层特征和深度特征分而治之, 最终两部分响应自适应融合得到最优结果. 文献[59]同样是从利用相

关滤波器的在线学习效率和 CNN 特征的判别能力进行结合的思路出发,在 ECO 框架的基础上提出了 CFWCR 方法. 首先从训练好的 CNN 的不同层提取的每个特征进行归一化,然后对每个特征的加权卷积响应进行求和,得到最终的响应得分. 文献[60]对 CFWCR 的工作进行了进一步的发展,提出了 MFT 算法. 针对 CNN 不同层次的特征和跟踪中遇到的不同问题之间的关联性,MFT 框架为不同的特性配置不同的特征,并且采用独立求解的方式,根据

视频的特性自适应调整融合权重,得到更加健壮的跟踪器. 文献[61]分析了深度特征的空间信息,提出了 LSART 跟踪方法. 首先,提出一种 KRR 模型,证明了该模型可以采用神经网络的方法进行求解. 然后,提出了一个具有空间正则化核的 CNN,每个输出通道对应的滤波器聚焦于目标的特定区域,并且通过距离变换池确定卷积层各输出通道的有效性. 最后,联合 KRR 模型和 CNN 的输出,得到最终的响应值.

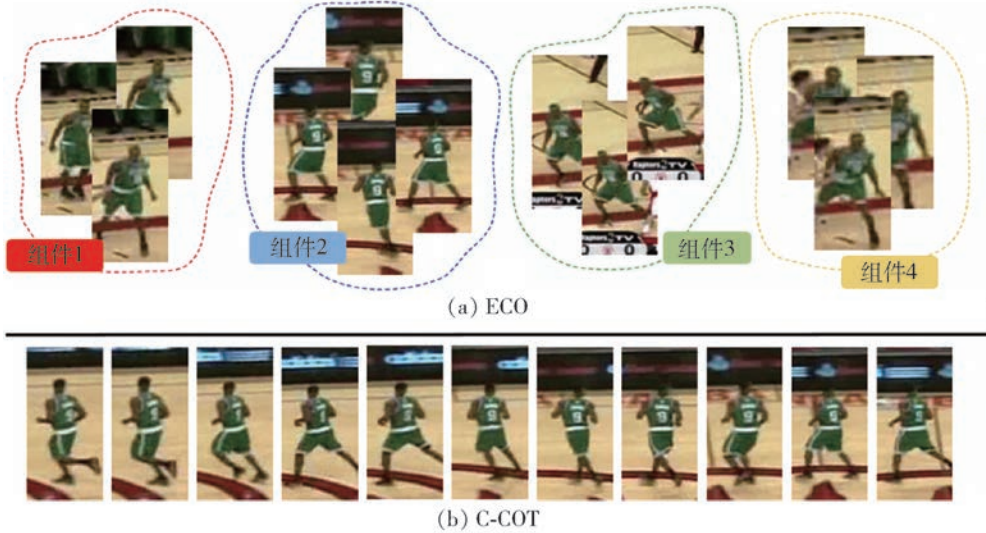


图 6 ECO 和 C-COT 的训练样本可视化示意图

Fig. 6 Visualization of the training set representation in the baseline ECO and C-COT

针对现有的空间正则化方法通过简单的预定义约束来正则化滤波器,忽略了特征输入的多样性和冗余性的缺陷,文献[62]提出了一种基于构建自适应空间特征选择外观模型的 LADCF 跟踪方法. 该框架利用时间一致性设计了一个低维判别流形空间,实现了可靠灵活的时间信息压缩,进而减少了滤波器的退化,保留了图像的多样性. 在实验时,只选择了 5% 的手工特征和 20% 的深度特征,却获得了更好的性能,同时解决了边界效应和背景杂波的问题. 为了更好地利用未标记的数据和样本空间的流形结构,文献[63]提出了一种基于增广样本的流形正则化 CFT-MRCT,利用矩阵的块循环结构从目标区域和非目标区域剪裁得到增广样本,通过标记和未标记的流形空间结构,并且采用半监督跟踪框架提升了跟踪性能. Mueller 等^[64]提出了一种基于内容注意力机制的相关滤波跟踪算法.

3.3 部件类 CFTs

与全局外观模型相比,许多跟踪器采用局部分

块策略建立外观模型. 其优势在于当目标发生部分遮挡时,未被遮挡的部分仍然可以有效地表示目标的外观模型,利用局部信息实现目标跟踪,而且分块模型也是实现尺度自适应的一种有效方式. 其劣势则体现为模型分块带来的计算复杂度问题,但是 CFTs 的高效性和鲁棒性可以弥补这一缺陷,使其应用于分块模型时具有极大的优势.

文献[33]提出了一种基于部件的实时跟踪方法 RPAC. 如图 7 所示,RPAC 将跟踪目标分解为 5 个局部目标,为每个部件配置独立的相关滤波器进行跟踪,然后将部件滤波器的置信得分联合得到最终的判定值. 为解决部件发生遮挡权重值仍较大引发误跟踪的问题,作者综合考虑峰值旁瓣比值与置信图平滑约束来判断部件的遮挡情况,并为每个部件分配权重,同时采用空间布局约束来抑制组合噪声的干扰. 文献[16]提出一种可信块跟踪方法,以 KCF 为基本滤波器,通过跟踪可信度度量来衡量块在跟踪时的可信程度,在序列蒙特卡罗框架下用概率模型估计可信块的分布,并通过霍夫投票来确定

目标的位置. 文献[65]通过发掘和利用可信的部件来学习目标模型,使用多方向递归神经网络遍历候选区域来捕捉上下文信息. 在跟踪过程中,利用神经网络生成置信度地图,在充分利用可靠部件信息的同时,通过正则化判别相关滤波来抑制背景噪声产生的影响. 文献[66]提出的 PKCF 算法同样以 KCF 为基滤波器训练每个模板块,通过粒子滤波框架自适应地设置每个块的权重.

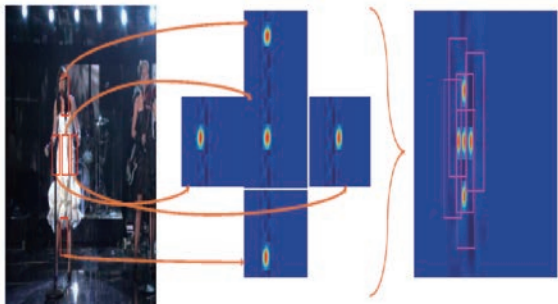


图7 RPAC 的算法框架

Fig. 7 Algorithm framework of RPAC

可变形部件在解决非刚性物体的形变和遮挡问题时显示出了巨大的潜力. 文献[34]提出的 DPCF 方法基于可变形部件相关滤波器进行目标跟踪,通过一个全局滤波器和多个部件滤波器间的耦合作用协同处理局部遮挡和尺度变化问题. 首先,通过局部滤波器对应特定的目标部件实现粗略估计,为全局滤波器提供定位依据. 然后,全局滤波器向部件滤波器提供反馈作为更新和相关形变参数的决策依据,具体算法框架如图8所示. 文献[67]提出的 DPT 方法也是一种基于相关滤波器的可变形部件跟踪方法,将视觉和几何特征约束为一个凸代价函数,并提出迭代直接法用于高效优化. 跟踪器在2个层次上对目标进行建模,综合了根相关滤波器作为粗略表示与中层星型部件模型用于精确定位,同时采取自顶向下定位、自底向上更新的方式实现目标跟踪. 文献[68]提出了一个统一的公式学习可变形卷积滤波器 DCCO. 在这个框架中,形变滤波器由子滤波器的线性组合进行表达,联合优化子滤波器的参数和它们之间的位置关系进行模型学习. 文献[69]使用马尔可夫随机模型捕捉块内时空关系和遮挡先验知识实现了跟踪性能的提升,并且使用一个高效密集的置信传播用于所提马尔可夫模型的推导.

文献[70]同样借鉴全局与局部相结合的思路,提出了 LGCF 视觉跟踪模型. 该模型有效联合了局部和全局策略,通过部件和目标之间的运动模型保

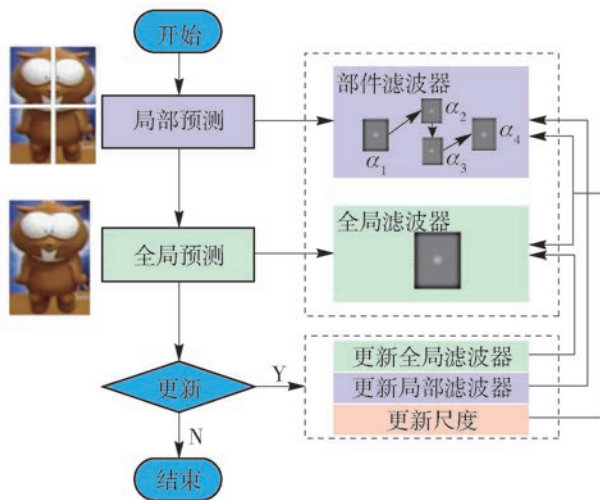


图8 DPCF 的算法框架

Fig. 8 Algorithm framework of DPCF

持其内部结构,同时加入时间一致性缓解了模型漂移的问题. 文献[71]提出一种利用部件空间结构的分块自适应核 CFT-SSPA-KCF,该模型同时最小化外观和形变代价以预测目标的位置. 优化过程分为2个部分,首先是采用 KCF 对部件的外观进行跟踪加快计算速度,然后通过结构化学学习机制最小化形变代价函数以减小标签噪声,最后结合最小生成树和动态规划得到的部件外观与形变得分确定目标在当前帧的位置.

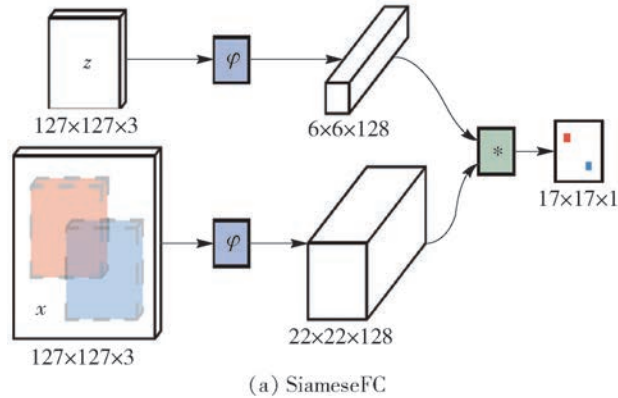
与目前将目标划分为固定数目的部件相比,文献[72]提出了一种基于动态可变形部件集合的非刚性目标跟踪方法. 其中,保留形状的核心滤波器被引入到水平集框架中用于动态地跟踪单个目标块,具备了假设复杂拓扑结构的能力. 当可变形部件捕获单个目标子区域时,使用光度判别和形状变化去显示单个目标子区域的跟踪性能,动态选择具有良好可跟踪性的子区域进行似然估计,最终实现目标轮廓的确定.

分块模型将局部图像特征与几何特征进行结合,是视觉跟踪的强大范例,具有一定的应对部分遮挡、目标形变和视角变化的能力. 其难点在于如何有效利用每个部件的时空置信图对全局目标位置进行估计,如何处理部件间的空间位置关系、全局目标与局部块间的关系,以及遮挡、形变情况下部件可信度的计算. 分块的数量和空间结构关系与跟踪的性能有着紧密联系,如何在密集搜索时兼备低计算成本是一个极具挑战的问题.

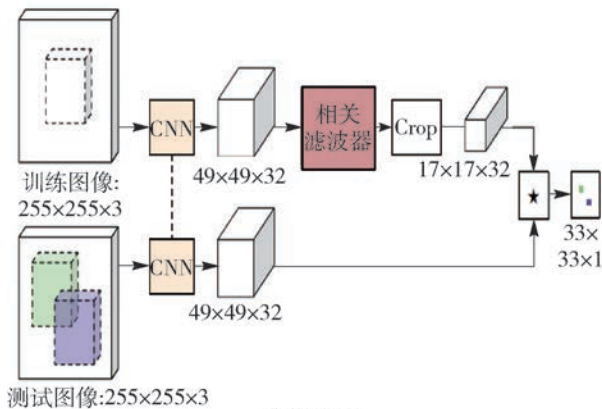
3.4 Siamese 网络类 CFTs

在最近的研究中,基于 Siamese 网络的视觉跟

踪算法因其高效性吸引了研究者的注意,并取得了优异的成绩. 1 个 Siamese 网络连接 2 个输入生成 1 个输出,通过共享卷积层或全连接层实现对 2 个图像块的相似性判别,目的是寻求不同输入图像中是否存在相同的目标. 文献 [73] 提出的 SiameseFC 框架通过目标区域和候选区域的相似性度量来完成跟踪任务,如图 9(a)所示, SiameseFC 是完全卷积的,输入为目标 z 和当前帧中大于目标的预测区域 x ,将 z 和 x 经过 φ 之后得到的输出送入相似性度量模块,最终输出结果为一个标量值响应分数地图,其尺寸取决于搜索图像的大小,分值最高的位置即为目标位置. 全卷积网络使得待搜索区域不需要与目标图像具有相同尺寸,为网络提供了更大的搜索区域,其中 φ 的体系结构与文献 [74] 中的类似,具体参数如表 1 所示. 为了应对跟踪过程中目标的尺度变化,搜索图像选取了 5 个尺度,分别是 $1.025^{-2}, -1.0, 1.2^1$. 前文中提到的文献 [24-25, 27] 显示了将深度特征与 CF 框架结合有助于提升跟踪器的性能,但是上述文献中的方法只是简单地利用提前训练好的深度特征,没有进行更加深入的整合. 针对这一局限性,并且考虑到 SiameseFC 全卷积结构缺少同类目标的判别性信息,文献 [26] 在



(a) SiameseFC



(b) CFNet

图 9 SiameseFC 和 CFNet 的算法框架

Fig. 9 Algorithm frameworks of SiameseFC and CFNet

SiameseFC 的基础上引入相关滤波器提出了 CFNet 算法,将具有封闭解的相关滤波器视为可微的 CNN 层,通过反向传播实现了端到端的学习. 如图 9(b) 给出的 CFNet 框架所示,在特征提取时,上下 2 个分支网络是相同的,差别体现在样例分支中加入的相关滤波器. 研究发现,CFNet 与 SiameseFC 相比,在网络层数较深时跟踪精度方面提升有限,性能并没有随着网络的加深而提高,甚至出现了降低,但是 CFNet 在网络层数较浅时可以取得较好的效果,也就是说 CFNet 能够得益于轻量级网络实现跟踪效率与准确性的兼顾. 不同于 CFNet 致力于改进 SiameseFC 的初衷,文献 [75] 同样提出了一种端到端的轻量级网络架构 DCFNet 框架,学习卷积特征并执行相关跟踪过程. 具体来说,DCFNet 使用卷积层在离线训练过程中编码先验信息,然后将 DCF 视为 Siamese 网络中增加的一个特殊的相关滤波层,通过它将网络输出定义为目标位置的概率热图,完成在线学习与跟踪. 虽然相关滤波层需要在线更新,但是因为推导是在傅里叶频域内进行的,所以得以保留了 DCF 的高效性.

表 1 SiameseFC 的网络参数

Table 1 Network parameters of SiameseFC

网络层	通道 地图	步长	样本 图像	搜索 图像	通道
			127 × 127	255 × 255	× 3
卷积层 1	96 × 3	2	59 × 59	123 × 123	× 96
池化层 1		2	29 × 29	61 × 61	× 96
卷积层 2	256 × 48	1	25 × 25	57 × 57	× 256
池化层 2		2	12 × 12	28 × 28	× 256
卷积层 3	384 × 256	1	10 × 10	26 × 26	× 192
卷积层 4	384 × 192	1	8 × 8	24 × 24	× 192
卷积层 5	256 × 192	1	6 × 6	22 × 22	× 128

作为匹配类跟踪算法中的一种代表性方法,孪生网络跟踪模型具有超越实时速度的计算效率,但是匹配类算法本身缺乏在线适应能力,难以捕捉跟踪过程中目标和背景发生的变化,很大程度上限制了跟踪的精确度. 针对这一缺陷,文献 [76] 在 SiameseFC 框架的基础上增加了目标外观变换转换层和背景抑制变换层,提出了动态孪生网络 Dsiam^[77] 跟踪方法. Dsiam 算法的主要贡献为: 1) 分别在模板分支和搜索分支引入外观变换转换层和背景抑制变换层,有效提升了模型的泛化能力. 2) 为了降低在线更新对计算效率产生的负面影响,并

且得益于模型的循环卷积形式,采用FFT对所提更新模型进行快速计算,兼顾了在线适应能力与高效性. 3) Dsiam实现了元素层面的多层特征融合,具备自适应融合不同层深度特征的能力. 文献[78]从平衡跟踪速度与精度的角度出发,提出了Early-Stopping跟踪器. 根据帧的难易程度自适应选择学习特征,利用马尔可夫决策求解跟踪问题,对易于处理的视频帧通过相关滤波结合人工特征完成,而对于困难帧采用孪生网络结合深度卷积特征完成,并且利用强化学习判断是否向后续卷积层继续,有效避免了需要最后一层做出决策的局限,提升了跟踪效率. 文献[79]提出的CREST算法将特征提取、生成响应地图和模型更新集成到神经网络中,采用end-to-end的训练方式,在跟踪过程中没有采用循

环移位和离散傅里叶变化,直接进行空间卷积运算,生成密集响应地图. 如图10给出的CREST算法框架所示,左边部分采用孪生网络提取特征地图,右边部分是在DCF Base mapping的基础上引入了2个残差学习:空间残差和时间残差. 其中,空间残差使用3层网络,输入为当前帧的特征地图,当目标受到如遮挡、光照变化等外部干扰因素的影响造成DCF无法回归到真值时,空间残差被用于进行响应修正. 时间残差的输入则是初始帧的特征地图,其作用是在空间残差与DCF均无法回归到真值时,对响应地图进行修正. 由此可见,残差学习的引入提高了外观发生剧烈变化时目标响应的鲁棒性.

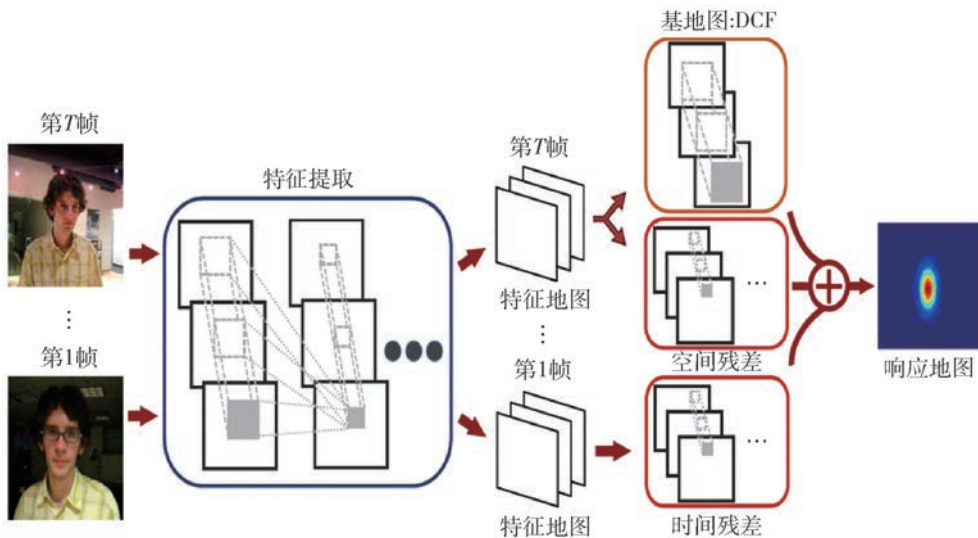


图10 CREST的算法框架

Fig. 10 Algorithm framework of CREST

结合深度卷积特征的相关滤波器在最近跟踪测试中取得了良好的成绩,但是目前大部分的跟踪器只考虑外观特征,而忽略了视频帧之间的运动信息. 针对这一问题,文献[80]提出了光流跟踪器FlowTrack. 与之前利用预先计算好光流信息用于跟踪不同,作者将特征提取、warp操作、特征融合、相关滤波器等模块集于一个网络并进行end-to-end训练. 具体来说,网络采用Siamese框架,分为历史分支和当前分支,在历史分支中进行光流的提取和warp,在当前分支只进行特征的提取. 在融合阶段,设计了一种时空注意力机制,在空间位置和时间维度上分别配置权重,并且采用类似SENet的方式度量权重. 完成特征融合之后,将历史分支和当前分支的输出一起送入CF层进行训练,得到最终的跟踪结果.

离线学习的本质使得Siamese网络无法真正意义上抑制背景中出现的干扰样本,也就是说很难正确区分2个外观类似的人或者车,导致跟踪发生漂移. 而相关滤波算法是在线学习的方法,可以通过分析上下文之间的关系进行调整,提升算法的性能. 因此,如何将在线学习与离线训练有效结合,是需要重点考虑的一个问题.

4 实验对比与分析

本节给出第3节研究现状分析中讨论的4种不同类别且具有代表性的一些相关滤波类跟踪算法在OTB2013^[81]、OTB100^[82]和VOT2017^[83]基准数据集上的性能对比与分析,3个基准数据集的详细信息如表2所示.

表 2 基准数据集的详细信息
Table 2 Details of different benchmarks

基准数据集	序列	最小帧数	最大帧数	总帧数	资源链接
OTB2013	50	71	3 872	29 491	http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html
OTB100	100	71	3 872	59 040	http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html
VOT2017	60	41	1 500	21 356	http://www.votchallenge.net/vot2017/

4.1 OTB 基准数据集

4.1.1 OTB2013 和 OTB100 数据集

OTB2013 基准数据集由 50 段带有标记属性的视频序列组成, 标记属性代表了现实场景中执行跟踪任务时会遇到的跟踪难题, 分别是光照变化 (illumination variation, IV)、尺度变化 (scale variation, SV)、遮挡 (occlusion, OCC)、形变 (deformation, DEF)、运动模糊 (motion blur, MB)、快

速运动 (fast motion, FM)、平面内旋转 (in-plane rotation, IPR)、平面外旋转 (out-of-plane rotation, OPR)、移出视野范围 (out-of-view, OV)、低分辨率 (low resolution, LR) 和背景杂乱 (background clutters, BC). 文献[82]在 OTB2013 的基础上对数据集进行了扩展, 将 50 段视频序列增加到了 100 段. OTB2013 基准数据集的视频序列显示如图 11 所示, 视频序列的属性标记情况如表 3 所示.

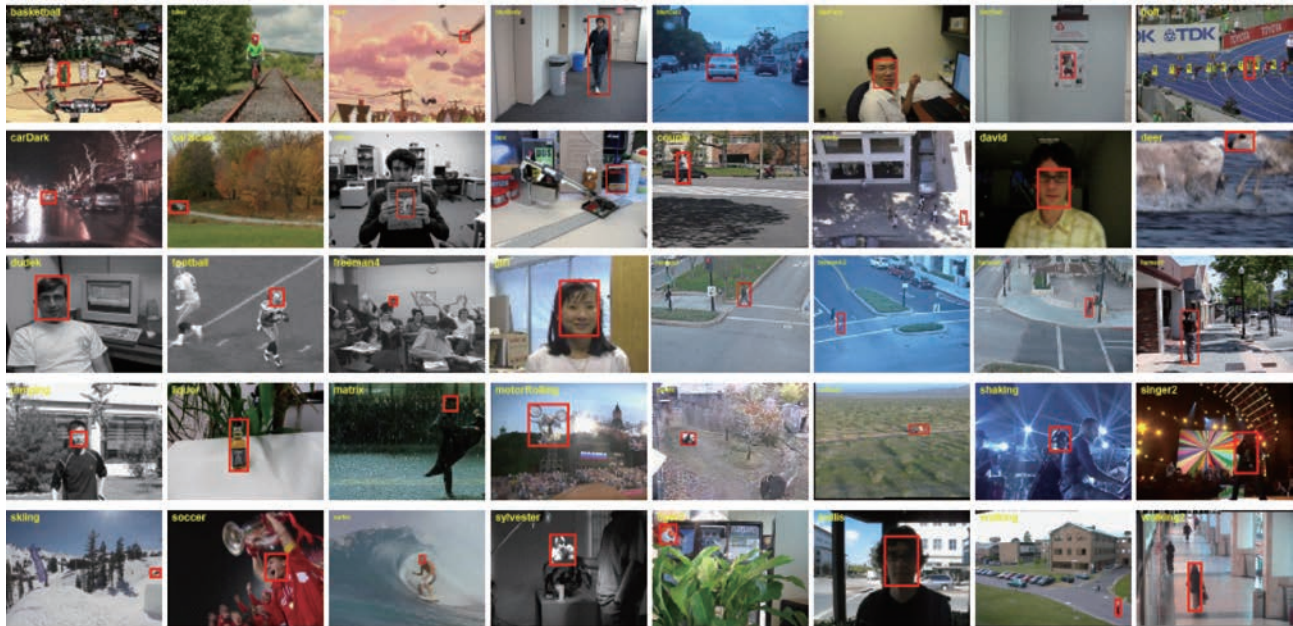


图 11 OTB2013 基准数据集的视频序列

Fig. 11 Video sequences of OTB2013 benchmark dataset

4.1.2 OTB 评价指标

为了评价跟踪器的性能, 文献[81]采用 2 种方式对其进行度量. 一种是精确度 (precision rate), 中心位置误差 (center location error, CLE) 是一种衡量跟踪精度的评价指标, 定义为被跟踪目标中心位置与地面真值之间的平均欧式距离. 另一种是成功率 (success rate), 若跟踪边界框与地面真值框之间的重叠率超过设定的阈值, 则视为跟踪成功. 成功率表示跟踪成功的帧数所占的比例, 重叠率计算公式为

$$S = \frac{|r_t \cap r_a|}{|r_t \cup r_a|} \quad (9)$$

式中: r_t 为跟踪框; r_a 为真值框; \cap 为两者重叠的区域; \cup 为两者的覆盖区域. 除此之外, 用精确度图 (precision plot) 和成功率图 (success plot) 评估跟踪器的性能. 精确度图显示了阈值在一定范围内的平均精度, 本节中阈值取 20. 成功率图表示阈值在一定范围内的重叠率精度, 由曲线下的面积评估跟踪器性能. 一般评估算法的方法是从第 1 帧的真值位置初始化跟踪器, 在整个测试序列中运行, 得到平均精度和成功率, 并定义为一次评估值 (one-pass evaluation, OPE). 然而, 跟踪器对初始化的敏感性导致不同的初始化导致性能间的巨大差异, 因此使

表3 OTB2013 视频序列的标记属性

Table 3 Tag attribute of the video sequences on OTB2013

序列名称	标记属性	序列名称	标记属性
Basketball	IV、OCC、DEF、OPR、BC	Biker	SV、OCC、MB、FM、OPR、OV
Bird1	FM、OV、DEF	BlurBody	SV、DEF、MB、FM、IPR
BlurFace	MB、FM、IPR	BlurOwL	SV、MB、FM、IPR
Bolt	OCC、DEF、IPR、OPR	Box	IV、SV、OCC、MB、IPR、OPR
Car1	IV、SV、MB、FM、BC、LR	Car4	IV、SV
CarDark	IV、BC	Carscale	SV、OCC、FM、IPR、OPR
ClifBar	SV、OCC、MB、FM、IPR、OV	BlurCar2	SV、MB、FM
Couple	SV、DEF、FM、OPR、BC	Crowds	IV、DEF、BC
David	IV、SV、OCC、DEF、MB、JPR、OPR	Deer	MB、FM、IPR、BC、LR
Diving	SV、DEF、IPR	Woman	IV、SV、OCC、DEF、MB、FM、OPR
DragonBaby	SV、OCC、MB、FM、IPR、OPR、OV	Dudek	SV、OCC、DEF、FM、IPR、OPR、OV
Football	OCC、IPR、OPR、BC	Freeman4	SV、OCC、IPR、OPR
Girl	SV、OCC、IPR、OPR	Human4	IV、SV、OCC、DEF
Human6	SV、OCC、DEF、FM、OPR、OV	Human9	IV、SV、DEF、MB、FM
Ironman	IV、SV、OCC、MB、FM、IPR、OPR	Jumping	MB、FM
Liquor	IV、SV、OCC、BM、FM、OPR、OV、BC	Jump	SV、OCC、DEF、MB、FM、IPR、OPR
Matrix	IV、SV、OCC、FM、IPR、OPR、BC	MotorRolling	IV、SV、MB、FM、IPR、BC、LR
Panda	SV、OCC、DEF、IPR、OPR、OV、LR	RedTeam	SV、OCC、IPR、OPR、LR
Shaking	IV、SV、IPR、OPR、BC	Singer2	IV、DEF、IPR、OPR、BC
Skating1	IV、SV、OCC、DEF、OPR、BC	Skating2	SV、OCC、DEF、FM、OPR
Skiing	IV、SV、DEF、IPR、OPR	Soccer	IV、SV、OCC、MB、FM、IPR、OPR
Surfer	SV、FM、IPR、OPR、LR	Sylvester	IV、IPR、OPR
Tiger2	IV、OCC、DEF、MB、FM、IPR、OPR、OV	Trellis	IV、SV、IPR、OPR、BC
Walking	SV、OCC、DEF	Walking2	SV、OCC、LR

用2种方法来分析跟踪器对初始化的鲁棒性,即对初始化分别进行时间和空间扰动,称为时间鲁棒性评价(temporal robustness evaluation, TRE)和空间鲁棒性评价(spatial robustness evaluation, SRE).

4.1.3 跟踪算法对比

为了全面而清晰地评估CFTs的性能,本节对32种具有代表性的CFTs在OTB2013和OTB100基准数据集上的测试结果进行了比较.根据第3节所提分类方式将32种CFTs归为4类.属于基本CFTs的有15种,包括KCF^[12]、KCF+MT^[30]、DSST^[17]、FDSST^[18]、SAMF^[20]、Staple^[23]、Staple-CA^[64]、KCFDPT^[40]、LMCF^[19]、SCT^[84]、LCT^[42]、SCF^[48]、

HDT^[38]、MCPF^[49]和MUSTer^[44];属于正则化CFTs的有10种,包括SRDCF^[51]、Deep-SRDCF^[25]、SRDCF-decon^[54]、C-COT^[27]、ECO^[28]、BACF^[85]、CSR-DCF^[52]、MRCT^[63]、SWCF^[86]和ACFN^[87];属于Siamese类CFTs的有5种,包括SiamFC^[73]、CFNet-conv1^[26]、CFNet-conv2^[26]、CREST^[79]和DCFNet^[75];属于部件类CFTs的有2种,包括DPCF^[34]和RPT^[16].上述方法的基本原理与跟踪框架已在第3章研究现状分析部分进行了分析,并且表4给出其中20种开源算法的资源链接.

之所以选择这32种CFTs进行对比,有以下4点原因:1)将本节与第3节研究现状分析部分进

表 4 20 种 CFTs 的资源链接
Table 4 Resource link of 20 types of CFTs

跟踪算法	类别	资源链接
CSRDCF	R-CFT	https://github.com/alanlukezic/csr-dcf
ECO	R-CFT	https://github.com/martin-danelljan/ECO
CCOT	R-CFT	https://github.com/martin-danelljan/Continuous-ConvOp
STRCF	R-CFT	https://github.com/lifeng9472/STRCF
SRDCF	R-CFT	https://www.cvl.isy.liu.se/research/objrec/visualtracking/regvistrack/
DeepSRDCF	R-CFT	https://www.cvl.isy.liu.se/research/objrec/visualtracking/regvistrack/
BACF	R-CFT	http://www.hamedkiani.com/bacf.html
DCFNet	S-CFT	https://github.com/foolwood/DCFNet#dcfnet-discriminant-correlation-filters-network-for-visual-tracking
CFNet	S-CFT	https://github.com/bertinetto/cfnet
SiameseFC	S-CFT	https://github.com/bertinetto/siamese-fc
HDT	P-CFT	https://sites.google.com/site/yuankiqi/hdt/
RPT	P-CFT	https://github.com/ihpdep/rpt
DPCF	P-CFT	http://www.votchallenge.net/vot2016/download/01_DPCF.zip
MCPF	B-CFT	http://nlpr-web.ia.ac.cn/mmc/homepage/tzzhang/mcpf.html
CF2	B-CFT	https://github.com/jbhuang0604/CF2
STAPLE	B-CFT	https://github.com/bertinetto/staple
FDSST	B-CFT	http://www.cvl.isy.liu.se/en/research/objrec/visualtracking/scalvistrack/index.html
DSST	B-CFT	http://www.cvl.isy.liu.se/en/research/objrec/visualtracking/scalvistrack/index.html
SAMF	B-CFT	https://github.com/ihpdep/samf
KCF	B-CFT	http://www.robots.ox.ac.uk/~joao/circulant/

行前后呼应,所选取的 32 种 CFTs 的设计思路与算法框架在第 3 节给出了介绍,通过理论知识与实验结果的结合加深对算法的理解;2) 所选 32 种算法涵盖了基本类、正则化类、部件类和 Siamese 网络类 CFTs,可以更加全面地了解 CFTs 的整体性能水平;3) 32 种 CFTs 中有 15 种属于基本类,并且涵盖了对 KCF 多个角度的改进,如特征表示、尺度变化的处理、核函数、长期跟踪、响应分布以及算法集成等;4) 同类算法之间具有较强的关联性,例如正则类中的 Deep-SRDCF、SRDCF-decon 和 C-COT 都是在 SRDCF 基础上的改进算法, Siamese 类中的 CFNet 和 DCFNet 都是在 SiameseFC 基础上提出的改进方案,通过对比这些具有一定关联性的算法有利于更深入地了解相关领域研究的发展现状及有价值的研究方向. 表 5 是对 32 种 CFTs 在 2 个基准数据集上的 2 个性能指标的统计结果,图 12 显示了它们在 OTB2013 和 OTB100 数据集上的成功率、精确率值及帧率. 根据给出的测试结果,对这些算法从特征的角度、模型和框架的角度、跟踪挑战的角度以及实

时性的角度分别进行对比分析.

从特征的角度进行分析,因为相关滤波类跟踪属于判别式跟踪方法,即将跟踪问题转换为检测问题,所以好的特征是此类跟踪的关键因素. 从表 5 给出的成功率和精确率结果以及表 6 给出的特征信息可以看出, HOG 和 CN 特征在视觉跟踪领域体现出了优异的性能,一系列基于 HOG 特征、CN 特征或者是两者融合的算法相继提出,如 RPT、DSST、FDSST、SAMF、KCFDPT、SRDCF 等. DeepSRDCF 在 SRDCF 的基础上,将深度卷积特征引入 SRDCF 中,探索了不同 conv 特征的性能差异,结果显示采用第 1 层 conv 的效果最好,在 2 个数据集上将成功率和精确率分别提升了 1.5%、3.7% 和 1.1%、6.2%. CFNet 同样对比了分别使用 conv1、conv2 和 conv5 时的性能差异,结果显示采用 conv2 的性能最优. 之后提出的许多方法以不同的方式结合深度特征构造跟踪框架,体现出了良好的性能,但是它们独立学习各层的 CFT 而忽视了层与层之间的关联性,使得此类方法存在一定的局限. CF2 和 HDT 从不同层

表5 32种CFTs在OTB2013和OTB100上的性能对比结果

Table 5 Performance comparison results of 32 types of CFTs on OTB2013 and OTB100

跟踪器	OTB2013		OTB100		跟踪器	OTB2013		OTB100	
	成功率	精确率	成功率	精确率		成功率	精确率	成功率	精确率
ECO	0.709	0.928	0.689	0.910	DPCF	0.605	0.829	0.565	0.776
BACF	0.678		0.630		CF2	0.605	0.891	0.562	0.837
MCPF	0.677	0.916	0.628	0.873	HDT	0.603	0.889	0.654	0.848
CREST	0.673	0.908	0.623	0.837	Staple	0.600	0.793	0.578	0.784
C-COT	0.672	0.899	0.671	0.896	CSR-DCF	0.599	0.800	0.598	0.733
SRDCF-decon	0.653	0.870	0.627	0.825	SCT	0.595	0.845		
MUSTer	0.641	0.865	0.575	0.774	KCFDPT	0.580	0.795	0.547	0.749
Deep-SRDCF	0.641	0.849	0.635	0.851	SAMF	0.579	0.785	0.541	0.752
LCT	0.628	0.848	0.562	0.762	RPT	0.577	0.803		
SRDCF	0.626	0.838	0.598	0.789	MRCT	0.565	0.805	0.507	0.736
LMCF	0.624	0.839	0.568		CFNet-conv1	0.545	0.753	0.516	0.691
SCF	0.623	0.874			KCF + MT	0.542	0.796	0.502	0.749
Staple-CA	0.621	0.833	0.598	0.81	FDSST	0.530	0.713	0.479	0.644
SiamFC	0.612	0.815			KCF	0.514	0.740	0.477	0.696
CFNet-conv2	0.611	0.807	0.568	0.748	DSST	0.503	0.737	0.470	0.693
ACFN	0.607	0.860	0.575	0.802	SWCF	0.491	0.662	0.454	0.622

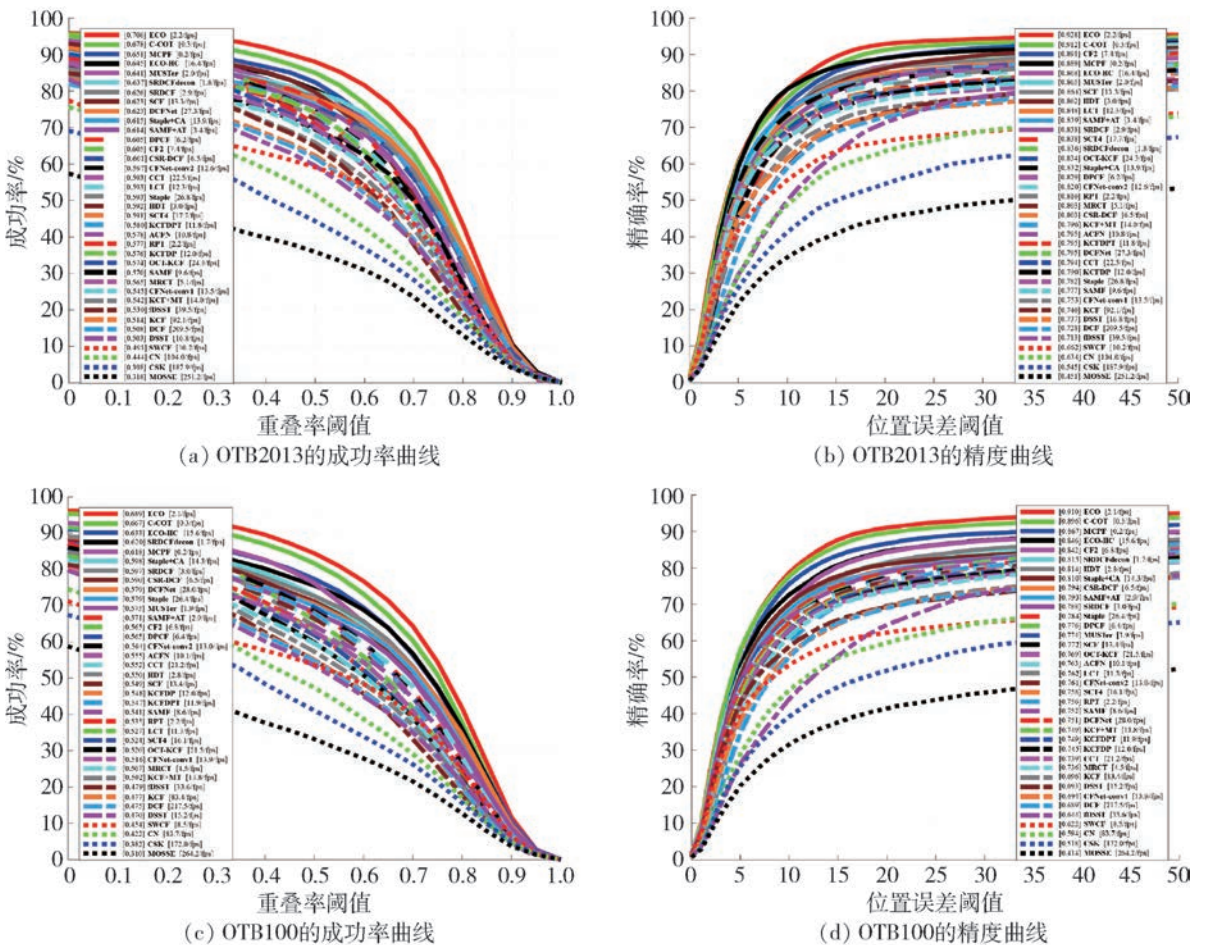


图12 相关滤波类跟踪器在OTB-2013和OTB100上的成功率曲线和精确率曲线

Fig. 12 Success plots and precision plots of CFTs on OTB-2013 and OTB100

表 6 CFTs 的详细信息

Table 6 Detailed information of selected trackers

跟踪算法	特征	尺度估计	离线训练	在线学习	跟踪算法	特征	尺度估计	离线训练	在线学习
KCF	Raw pixels、HOG	否	否	是	SRDCF	HOG、CN	否	否	是
KCF + MT	HOG	是	否	是	C-COT	CNN	是	否	是
SAMF	Raw pixels、HOG、CN	是	否	是	ECO	CNN、HOG、CN	是	否	是
DSST	HOG	是	否	是	CSRDCF	HOG、CN、Color	是	否	是
FDSST	HOG	是	否	是	BACF	HOG	是	否	是
CF2	CNN	否	否	是	MRCT	HOG、Intensity、IIF	否	否	是
HDT	CNN	否	否	是	SWCF	HOG、Intensity	是	否	是
KCFDPT	HOG、CN	是	否	是	ACFN	HOG、Color	是	否	是
LMCF	HOG、CN	是	否	是	SiameseFC	CNN	是	是	否
STAPLE	HOG、Color Histogram	是	否	是	CFNet	CNN	是	是	是
LCT	HOG	是	否	是	DCFNet	CNN	是	是	是
MCPF	CNN	是	否	是	CREST	CNN	是	是	是
MUSTer	HOG	否	否	是	RPT	HOG	否	否	是
SCT	RGB、HOG、Lab Color	否	否	是	DKCF	HOG、Color	是	否	是
SCF	HOG、CN	是	否	是					

特征优势互补的角度出发,在不影响实时性的同时提升了跟踪性能,MCPF在其基础上进一步利用不同层之间的依赖关系训练滤波器,进一步提升了跟踪精度. C-COT使用VGG-Net提取深度特征,将原始的彩色图像和2个卷积层的输出作为特征,与CFNet等同类算法相比精确度有了显著提升,但是复杂的特征严重降低了计算效率,难以达到实时性的要求. ECO在C-COT基础上通过因式分解操作,将HOG、CN和CNN的特征维度进行降维,其中HOG压缩到10,CN压缩到3,CNN的第1、5卷积层分别压缩到16和64,减少了训练参数,进而有效降低了计算复杂度,使得帧率提高了大约8倍.

根据表5的对比结果结合表6的跟踪器详细信息对32种CFTs从模型和框架的角度进行分析,DSST在KCF基础上从解决尺度问题的角度出发,扩展了跟踪框架,增加了一个独立的尺度滤波器,与KCF相比,准确率有了一定的提高.之后提出的加速版FDSST采用降维操作和QR分解降低了计算量,帧率与DSST相比提高了1倍多,同时在OTB2013上的成功率也提升了近3%.SAMF与DSST算法类似,同样采用了尺度穷举的思路,但是尺度因子的变化范围较小,不能较好地处理尺度变化较大的情况.与尺度

池方法不同,DPCF基于可变形分块实现了对滤波器大小的自适应调节,在OTB2013和OTB100上的成功率和精确率与SAMF相比分别提升了2.6%、4.4%和2.4%、2.4%.针对判别相关滤波器采用周期假设导致不必要的边界效应这一问题,SRDCF引入空间正则化弱化背景信息的干扰,显著提升了性能. SRDCF-decon在SRDCF的基础上从提升样本质量的角度出发,在OTB2013上的成功率提升了3.3%.CSR-DCF利用空间置信图使滤波器参数同样是集中在置信分高的地方,和SRDCF一样抑制背景信息的干扰,但是采取的方式不同,性能差距也较大. BACF同样对边界效应进行了处理,提出了扩大所搜区域采用真实负样本的思路,取得了比SRDCF-decon更加优异的成绩. BACF的成绩很大程度上说明了样本质量对跟踪器性能有很大的影响.不同类跟踪算法根据被跟踪目标和任务的不同设计算法框架,均具有自身的优势及劣势,所以将相关滤波器与其他算法集合去实现优势互补也是一种提升跟踪性能的思路. LMCF通过集成CF与结构化SVM,与KCF相比,性能得到了显著提升. MCPF跟踪框架将MCF与粒子滤波器进行了结合,采用粒子抽样策略有效处理了尺度变化问题,并引导抽样粒子向目标状态分布的模式方向运动提

升了跟踪性能. 近期一些将 Siamese 网络与相关滤波结合的算法也被用于处理跟踪问题, 如 SiamFC、CFNet 和 DCFNet 等, 图 12 给出的帧率说明了使用轻量级网络能够实现跟踪速度方面的提升.

从跟踪挑战的角度进行分析, 大多数跟踪器并不能在所有的跟踪挑战中都表现出良好的性能. OTB 数据集将跟踪挑战的属性归纳为分为 IV、SV、OCC、DEF、MB、FM、IPR、OPR、OV、LR 和 BC, 属性的具体介绍已在 4.1.1 小节给出. 表 7、8 分别给出了 10 种采用手工特征的 CFTs 和 10 种采用深度特征的 CFTs 在 OTB100 基准数据集不同挑战属性下得到的 Success Rate, 排名第一的结果用粗体表示, 排名第二的结果用斜体表示, 排名第三的结果用下划线表示. 在快速运动和运动模糊挑战中, 目标的外观因目标自身或者是相机的运动而模糊. 手工特征 CFTs 中 ECO-HC 和 SRDCF-decon 表现最好, 这是因为 SRDCF-decon 通过扩大搜索空间的方式解决了快速运动序列中目标位置发生较大变化的难题, 而 ECO-HC 采用亚像素级的多分辨率特征有效提升了跟踪精度. 在深度特征 CFTs 中 ECO 和 C-COT 表现最好. 在尺度变化和遮挡挑战中, 跟踪器在目标观测时通常会将干扰背景作为目标信息而导致跟踪漂移的发生. 在手工特征 CFTs 中, SRDCF、CSRDCF 和 ECO-HC 表现最好, 而深度特征 CFTs 中 ECO、C-COT 和 MCPF 表现最好. 综合性能最好的 ECO 通过混合高斯模型对训练样本进行分组, 不仅避免了近期错误样本产生的过拟合问题, 而且提升

了样本的多样性与判别能力, 因此表现出了优异的性能. 从表 3 给出的 OTB 视频序列的标记属性可以看出, 在许多测试序列中, 光照变化和低分辨率是同时存在的, 为准确跟踪带来了很大的难题. 为了解决光照问题, 跟踪器需要提升对目标模型和背景模型的判别能力, 而在低分辨率图像中, 外观表示也起到了重要作用. 因此, 这 2 个挑战可以通过表征能力强的特征解决. 在手工特征 CFTs 中, ECO-HC 和 SRDCF-decon 表现最好, 在深度特征 CFTs 中, ECO 和 CF-Net 在精度方面表现最好, 并且深度特征与手工特征相比具有更强的表征能力. 移出视野范围对于大多数跟踪器而言是最为严峻的挑战之一, 通过保持有用的目标样本在跟踪失败后启动重检测模块是解决 OV 问题的途径. 手工特征 CFTs 中, CSRDCF 和 SRDCF 具有较高的跟踪精度, 而采用深度特征的 ECO 和 C-COT 得益于学习多分辨率特征有效应对了 OV 挑战. 由于判别型跟踪方法在线学习正负样本, 背景杂乱会对跟踪器的训练造成干扰, 因此背景杂乱也是难以处理的跟踪挑战之一, ECO 和 ECO-HC 表现出了优异的性能. 因为 ECO 采用多分辨率卷积特征提升了外观表示, 利用高斯混合模型对样本集进行分组提升了样本质量, 稀疏型模型更新方式提升了跟踪效率, 所以在绝大多数跟踪挑战中取得了最为优异的性能. 通过上述分析可以看出, 4 种类别的 CFTs 相比, 每个挑战属性下表现最为良好的都是正则类 CFTs.

表 7 手工特征 CFTs 在 OTB100 不同属性测试序列中的成功率

Table 7 Success rate of handcrafted feature trackers on OTB100

属性	ECO-HC	SRDCF	CSR-DCF	SRDCF-decon	DPCF	STAPLE	SAMF	FDSST	RPT	KCF
IV	0.631	<u>0.609</u>	0.575	<i>0.628</i>	0.585	0.595	0.533	0.476	0.537	0.479
SV	0.603	<u>0.561</u>	0.531	<i>0.600</i>	0.517	0.520	0.482	0.398	0.481	0.394
OCC	0.620	<i>0.578</i>	0.556	<i>0.578</i>	0.546	0.543	0.543	0.418	0.480	0.443
DEF	0.573	0.544	<i>0.562</i>	0.549	0.523	<u>0.550</u>	0.516	0.414	0.491	0.436
MB	<i>0.611</i>	0.594	<u>0.610</u>	0.626	0.573	0.540	0.534	0.495	0.517	0.459
FM	0.636	<u>0.597</u>	0.592	<i>0.602</i>	0.533	0.541	0.525	0.490	0.542	0.459
IPR	0.574	0.541	0.540	<i>0.563</i>	0.525	<u>0.549</u>	0.503	0.486	0.523	0.469
OPR	0.601	<u>0.548</u>	0.541	<i>0.581</i>	0.546	0.534	0.528	0.436	0.503	0.453
OV	0.600	0.461	<u>0.517</u>	<i>0.543</i>	0.482	0.476	0.483	0.397	0.471	0.393
LR	<i>0.534</i>	<u>0.514</u>	0.460	0.535	0.409	0.399	0.348	0.295	0.371	0.290
BC	0.619	<u>0.583</u>	0.565	<i>0.615</i>	0.575	0.561	0.510	0.527	0.575	0.498

表 8 深度特征 CFTs 在 OTB100 不同属性测试序列中的成功率

Table 8 Success rate of deep feature trackers on OTB100

属性	ECO	CCOT	MCPF	CFNet-conv1	CFNet-conv2	DeepSRDCF	CF2	HDT	ACFN	DCFNet
IV	0.711	0.672	<u>0.611</u>	0.478	0.532	0.603	0.549	0.528	0.545	0.581
SV	0.664	0.653	<u>0.576</u>	0.476	0.519	0.564	0.485	0.470	0.526	0.564
OCC	0.676	0.656	<u>0.601</u>	0.477	0.524	0.562	0.532	0.512	0.503	0.573
DEF	0.635	0.608	<u>0.577</u>	0.460	0.475	0.559	0.530	0.529	0.524	0.494
MB	0.705	0.694	<u>0.624</u>	0.516	0.577	0.541	0.585	0.560	0.531	0.544
FM	0.680	0.670	<u>0.580</u>	0.495	0.552	0.546	0.570	0.554	0.540	0.543
IPR	0.651	0.620	<u>0.606</u>	0.511	0.575	0.541	0.566	0.542	0.530	0.556
OPR	0.672	0.643	<u>0.604</u>	0.479	0.540	0.548	0.540	0.525	0.515	0.573
OV	0.651	0.623	0.532	0.362	0.423	0.513	0.474	0.453	0.457	<u>0.558</u>
LR	<u>0.590</u>	0.608	0.502	0.609	0.576	0.557	0.388	0.402	0.491	0.504
BC	0.699	0.657	<u>0.610</u>	0.444	0.567	0.559	0.585	0.553	0.520	0.569

从实时性的角度进行分析,通常情况下认为跟踪算法的帧率达到 25 帧/s 视为满足实时处理的需求. 从表 5 给出的实时性指标可以看出所对比的 32 种 CFTs 中,有 12 种可以实现实时跟踪. 最初提出的 MOSSE、CSK、CN 等算法的帧率均超过了 100 帧/s,之后提出的相关滤波类跟踪器从不同的优化角度出发有效提升了整体性能,但是越来越复杂的框架难免在一定程度上影响运算效率,即便如此,大部分的 CFTs 仍能满足或接近满足实时性的需求.

4.2 VOT 基准数据集

4.2.1 VOT2017 基准数据集

VOT2017^[88] 基准数据集由 60 段带有 5 种标记属性的视频序列组成,这 5 种属性分别是相机移动、光照变化、尺度变化、动作变化和遮挡. VOT2017 将 VOT2016 中已经被大多数算法准确跟踪的 10 个序列进行了替换,并且保持了序列的分布属性. VOT 数据集对于极小目标跟踪是非常重要的一个基准数据集,同时 VOT2017 对所有序列的真值进行了重新标定,精确到像素级别,然后重新拟合矩形框,给出了更为准确的标注信息.

4.2.2 VOT2017 评价指标

VOT2017 使用平均重叠期望 (expected average overlap, EAO)、Accuracy 和 Robustness 这 3 个性能指标对算法进行评估. Accuracy 指的是跟踪器在单个测试序列下真值与预测包围框之间平均重叠率,

重叠率的定义已在 4.1.2 小节给出. Robustness 指的是跟踪器在单个测试序列中失败的次数,当重叠率为 0 时即判定为跟踪失败,在失败并丢弃 5 帧之后进行重新初始化. EAO 是对每个跟踪器在一段序列上非重置重叠的期望值.

4.2.3 跟踪结果对比

为了进一步分析 CFTs 的性能,本小节对 10 种具有代表性的 CFTs 在 VOT2017 基准数据集上进行了比较. 10 种 CFTs 中包含 5 种使用深度特征的方法,分别为 CF2、ECO、C-COT、SiameseFC 和 MCPF. 另外 5 种使用人工特征的方法,为 CSRDCF、SRDCF、STAPLE、DSST 和 KCF. 其中 CSRDCF、SRDCF、C-COT 和 ECO 属于正则类,STAPLE、DSST、CF2 和 MCPF 属于基本类, SiameseFC 属于孪生网络类. 表 9 是这 10 种 CFTs 在 VOT2017 基准数据集上对 3 个性能指标的评估结果,得分最优的用粗体表示.

从表 9 给出的实验结果可以看出,在 Baseline 测试中,手工特征跟踪器中 CSRDCF 获得 EAO 和 R 的最优值分别为 0.256 和 0.356, STAPLE 获得 A 的最优值为 0.530. 深度特征跟踪器中 CF2 获得为 EAO 的最高得分为 0.286, MCPF 获得 A 的最高得分为 0.510, ECO 获得 R 的最高得分为 0.276. 在 Realtime 测试中,手工特征跟踪器中的 STAPLE 在各项指标中表现都是最好,深度特征跟踪器则由 SiameseFC 获得了各项指标的最高得分.

表9 HC特征和深度特征跟踪器在VOT2017基线和实时测试中的A、R和EAO对比结果

Table 9 Comparison of HC and deep CFTs using A, R and EAO for baseline and realtime experiments over VOT2017

跟踪算法	性能指标	基线			实时		
		EAO	A	R	EAO	A	R
手工特征 CFTs	CSRDCF	0.256	0.491	0.356	0.099	0.477	1.054
	STAPLE	0.169	0.530	0.688	0.170	0.530	0.688
	SRDCF	0.119	0.490	0.974	0.058	0.377	1.999
	DSST	0.079	0.395	1.452	0.077	0.396	1.480
	KCF	0.135	0.447	0.773	0.134	0.445	0.782
深度特征 CFTs	CF2	0.286	0.509	0.281	0.059	0.339	1.723
	ECO	0.280	0.483	0.276	0.078	0.449	1.466
	CCOT	0.267	0.494	0.318	0.058	0.326	1.461
	SiameseFC	0.188	0.502	0.585	0.182	0.502	0.504
	MCPF	0.248	0.510	0.427	0.060	0.325	1.489

5 结论

本文针对相关滤波类目标跟踪方法应用于实际场景时面临的主要难题,以及目前该领域的研究现状展开讨论与分析,并将32种代表性相关滤波类跟踪器在OTB-2013、OTB100和VOT2017基准数据集上进行了评估,表明了相关滤波类跟踪器在性能方面的优越性。虽然相关滤波类跟踪方法与其他方法相比具有一定的优势,但是在应用于实际场景时,准确性和鲁棒性仍然难以满足跟踪任务的需要。因此,目标跟踪领域以及相关滤波类跟踪仍然存在进一步深入研究与提升的空间。

结合文章的研究现状分析和实验部分的对比结果,并且考虑目前机器视觉领域的研究热点,目标跟踪的进一步研究可以从以下几个方面考虑:

1) 如何平衡跟踪性能与实时性之间的关系。复杂特征以及算法框架在提升跟踪精度的同时将不可避免地影响其实时性,所以做到两者兼顾才能实现高效稳健的跟踪。

2) 对于视觉跟踪任务,特征的优劣是影响跟踪性能的关键因素。在处理跟踪问题时,卷积特征与人工特征相比具有一定的优势,但是选取预先训练好的卷积特征还是跟踪专用的特征,以及何种网络训练得到的特征更利于跟踪仍是目前需要进一步研究探讨的课题。

3) 目前大部分单目标跟踪算法的研究对象都是短时目标跟踪任务,虽然已有少数研究者对长时

跟踪进行了研究,但其本质依然是在短时跟踪框架中引入重检测模块或其他校正模块来实现,缺乏具有针对性的长时跟踪框架,这是值得进一步深入研究的问题。

4) 相关滤波类跟踪算法属于一种 tracking-by-detection 框架,在线更新的学习方式和有限的训练样本可能导致过拟合问题,造成跟踪失败。而生成对抗网络(generative adversarial network, GAN)具有生成逼近真实图像的能力,在相关滤波跟踪框架中加入GAN用于生成训练样本是未来研究的一个重要方向。

5) 定制网络架构。探索更加适用于跟踪框架学习的网络架构,而不仅仅是通过迁移学习的方式进行网络微调,这同样是未来跟踪任务一个有前途的研究方向。

参考文献:

- [1] BROWN M, FUNKE J, ERLIEN S, et al. Safe driving envelopes for path tracking in autonomous vehicles [J]. Control Engineering Practice, 2017, 61: 307-316.
- [2] TIAN B, YAO Q, GU Y, et al. Video processing techniques for traffic flow monitoring: a survey [C] // Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on. Piscataway: IEEE, 2011: 1103-1108.
- [3] SEVERSON J. Human-digital media interaction tracking: US2019282154A1[P]. 2019-09-19.

- [4] ONATE J M B, CHIPANTASI D J M, ERAZO N D R V. Tracking objects using artificial neural networks and wireless connection for robotics [J]. *Journal of Telecommunication, Electronic and Computer Engineering*, 2017, 39(8): 161-164.
- [5] WALKER S, SEWELL C, PARK J, et al. Systems and methods for localizing, tracking and/or controlling medical instruments: US 2017209224AI[P]. 2017-07-27.
- [6] AGGARWAL J K, XIA L. Human activity recognition from 3d data: a review[J]. *Pattern Recognition Letters*, 2014, 48(1): 70-80.
- [7] YILMAZ A, JAVED O, SHAH M. Object tracking: a survey[J]. *ACM Computing Surveys (CSUR)*, 2006, 38(4): 1-45.
- [8] CANNONS K. A review of visual tracking[R]. Toronto: York University, 2008.
- [9] SMEULDERS A W, CHU D M, CUCCHIARA R, et al. Visual tracking: an experimental survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(7): 1442-1468.
- [10] LI X, HU W, SHEN C, et al. A survey of appearance models in visual object tracking[J]. *ACM Transactions on Intelligent Systems and Technology*, 2013, 4(4): 1884-2021.
- [11] YANG H, SHAO L, ZHENG F, et al. Recent advances and trends in visual tracking: a review [J]. *Neurocomputing*, 2011, 74(18): 3823-3831.
- [12] 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.
- [13] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2010: 2544-2550.
- [14] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C] // *European Conference on Computer Vision*. Berlin: Springer, 2012: 702-715.
- [15] DANELLJAN M, SHAHBAZ KHAN F, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2014: 1090-1097.
- [16] LI Y, ZHU J, HOI S C H. Reliable patch trackers: robust visual tracking by exploiting reliable patches [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2015: 353-361.
- [17] DANELLJAN M, HAGER G, KHAN F, et al. Accurate scale estimation for robust visual tracking [C] // *British Machine Vision Conference*, Nottingham, September 1-5, 2014. Leeds: BMVA, 2014: 1-11.
- [18] DANELLJAN M, HAGER G, KHAN F S, et al. Discriminative scale space tracking [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(8): 1561-1575.
- [19] WANG M, LIU Y, HUANG Z. Large margin object tracking with circulant feature maps [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2017: 4021-4029.
- [20] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration [C] // *European Conference on Computer Vision*. Berlin: Springer, 2014: 254-265.
- [21] CAMPLANI M, HANNUNA S, MIRMEHDI M, et al. Real-time RGB-D tracking with depth scaling kernelised correlation filters and occlusion handling [C] // *British Machine Vision Conference 2015*. Leeds: BMVA, 2015: 1-11.
- [22] ZHU G, WANG J, WU Y, et al. MC-HOG correlation tracking with saliency proposal [C] // *Thirtieth AAAI Conference on Artificial Intelligence*. Palo Alto: AAAI, 2016: 3690-3696.
- [23] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2016: 1401-1409.
- [24] MA C, HUANG J B, YANG X, et al. Hierarchical convolutional features for visual tracking [C] // *Proceedings of the IEEE International Conference on Computer Vision*. Piscataway: IEEE, 2015: 3074-3082.
- [25] DANELLJAN M, HAGER G, SHAHBAZ KHAN F, et al. Convolutional features for correlation filter based visual tracking [C] // *Proceedings of the IEEE International Conference on Computer Vision Workshops*. Piscataway: IEEE, 2015: 58-66.
- [26] VALMADRE J, BERTINETTO L, HENRIQUES J, et al. End-to-end representation learning for correlation filter based tracking [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE, 2017: 2805-2813.
- [27] DANELLJAN M, ROBINSON A, SHAHBAZ KHAN F, et al. Beyond correlation filters: learning continuous

- convolution operators for visual tracking [C] // European Conference on Computer Vision. Berlin: Springer, 2016: 472-488.
- [28] DANELLJAN M, BHAT G, SHAHBAZ KHAN F, et al. Eco: efficient convolution operators for tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6638-6646.
- [29] BHAT G, JOHNDER J, DANELLJAN M, et al. Unveiling the power of deep tracking [C] // Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 483-498.
- [30] BIBI A, GHANEM B. Multi-template scale-adaptive kernelized correlation filters [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2015: 50-57.
- [31] LI J, HONG Z, ZHAO B. Robust visual tracking by exploiting the historical tracker snapshots [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2015: 41-49.
- [32] SOLIS M A, LANG J, LAGASIERE R. Scalable kernel correlation filter with sparse feature integration [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2015: 24-31.
- [33] LIU T, WANG G, YANG Q. Real-time part-based visual tracking via adaptive correlation filters [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 4902-4912.
- [34] AKIN O, ERDEM E, ERDEM A, et al. Deformable part-based tracking by coupled global and local correlation filters [J]. Journal of Visual Communication and Image Representation, 2016, 38: 763-774.
- [35] ZHANG W, KANG B S. Recent advances in correlation filter-based object tracking: a review [J]. Journal of Image and Graphics, 2017, 22(8): 1017-1033.
- [36] MA C, XU Y, NI B, et al. When correlation filters meet convolutional neural networks for visual tracking [J]. IEEE Signal Processing Letters, 2016, 23(10): 1454-1458.
- [37] MA C, HUANG J B, YANG X, et al. Robust visual tracking via hierarchical convolutional features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 41(11): 2709-2723.
- [38] QI Y, ZHANG S, QIN L, et al. Hedged deep tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4303-4311.
- [39] ZHANG M, XING J, GAO J, et al. Joint scale-spatial correlation tracking with adaptive rotation estimation [C] // Proceedings of the IEEE International Conference on Computer Vision Workshops. Piscataway: IEEE, 2015: 32-40.
- [40] HUANG D, LUO L, WEN M, et al. Enable scale and aspect ratio adaptability in visual tracking with detection proposals [J]. Proceedings of the BMVC, 2015, 185: 1-12.
- [41] TANG M, FENG J. Multi-kernel correlation filter for visual tracking [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 3038-3046.
- [42] MA C, YANG X, ZHANG C, et al. Long-term correlation tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 5388-5396.
- [43] MA C, HUANG J B, YANG X, et al. Adaptive correlation filters with long-term and short-term memory for object tracking [J]. International Journal of Computer Vision, 2018, 126(8): 771-796.
- [44] HONG Z, CHEN Z, WANG C, et al. Multi-store tracker (muster): a cognitive psychology inspired approach to object tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 749-758.
- [45] BIBI A, MUELLER M, GHANEM B. Target response adaptation for correlation filter tracking [C] // European Conference on Computer Vision. Berlin: Springer, 2016: 419-433.
- [46] SUI Y, ZHANG Z, WANG G, et al. Real-time visual tracking: promoting the robustness of correlation filter learning [C] // European Conference on Computer Vision. Berlin: Springer, 2016: 662-678.
- [47] ZHANG B, LI Z, CAO X, et al. Output constraint transfer for kernelized correlation filter in tracking [J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2017, 47(4): 693-703.
- [48] ZUO W, WU X, LIN L, et al. Learning support correlation filters for visual tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 45(5): 1158-1172.
- [49] ZHANG T, XU C, YANG M H. Multi-task correlation particle filter for robust object tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 4335-4343.
- [50] FAN H, LING H. Parallel tracking and verifying: a

- framework for real-time and high accuracy visual tracking [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 5486-5494.
- [51] DANELLJAN M, HAGER G, SHAHBAZ K F, et al. Learning spatially regularized correlation filters for visual tracking [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 4310-4318.
- [52] LUKEŽIČ A, VOJIŘ T, Č EHOVIN ZAJC L, et al. Discriminative correlation filter with channel and spatial reliability [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 6309-6318.
- [53] DU F, LIU P, ZHAO W, et al. Spatial-temporal adaptive feature weighted correlation filter for visual tracking [J]. Signal Processing: Image Communication, 2018, 67: 58-70.
- [54] DANELLJAN M, HAGER G, SHAHBAZ K F, et al. Adaptive decontamination of the training set: a unified formulation for discriminative visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1430-1438.
- [55] GLADH S, DANELLJAN M, KHAN F S, et al. Deep motion features for visual tracking [C] // 2016 23rd International Conference on Pattern Recognition (ICPR). Piscataway: IEEE, 2016: 1243-1248.
- [56] DANELLJAN M, BHAT G, GLADH S, et al. Deep motion and appearance cues for visual tracking [J]. Pattern Recognition Letters, 2019, 124: 74-81.
- [57] LI F, TIAN C, ZUO W, et al. Learning spatial-temporal regularized correlation filters for visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4904-4913.
- [58] FAN J, SONG H, ZHANG K, et al. Complementary tracking via dual color clustering and spatio-temporal regularized correlation learning [J]. IEEE Access, 2018, 6: 56526-56538.
- [59] HE Z, FAN Y, ZHUANG J, et al. Correlation filters with weighted convolution responses [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1992-2000.
- [60] BAI S, HE Z, XU T B, et al. Multi-hierarchical independent correlation filters for visual tracking [J]. arXiv Preprint arXiv, 2018: 1811. 10302.
- [61] SUN C, WANG D, LU H, et al. Learning spatial-aware regressions for visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8962-8970.
- [62] XU T, FENG Z H, WU X J, et al. Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual tracking [J]. arXiv Preprint arXiv, 2018: 1807. 11348.
- [63] HU H, MA B, SHEN J, et al. Manifold regularized correlation object tracking [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(5): 1786-1795.
- [64] MUELLER M, SMITH N, GHANEM B. Context-aware correlation filter tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 1396-1404.
- [65] CUI Z, XIAO S, FENG J, et al. Recurrently target-attending tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1449-1458.
- [66] CHEN W, ZHANG K, LIU Q. Robust visual tracking via patch based kernel correlation filters with adaptive multiple feature ensemble [J]. Neurocomputing, 2016, 214: 607-617.
- [67] LUKEZIC A, ZAJC L Č, KRISTAN M. Deformable parts correlation filters for robust visual tracking [J]. IEEE Transactions on Cybernetics, 2018, 48(6): 1849-1861.
- [68] JOHNDER J, DANELLJAN M, KHAN F S, et al. DCCO: towards deformable continuous convolution operators for visual tracking [C] // International Conference on Computer Analysis of Images and Patterns. Berlin: Springer, 2017: 55-67.
- [69] YAO R, XIA S, ZHANG Z, et al. Real-time correlation filter tracking by efficient dense belief propagation with structure preserving [J]. IEEE Transactions on Multimedia, 2017, 19(4): 772-784.
- [70] FAN H, XIANG J. Robust visual tracking via local-global correlation filter [C] // AAAI Conference on Artificial Intelligence. Palo Alto: AAAI, 2017: 4025-4031.
- [71] YAO R, XIA S, SHEN F, et al. Exploiting spatial structure from parts for adaptive kernelized correlation filter tracker [J]. IEEE Signal Process Lett, 2016, 23(5): 658-662.
- [72] SUN X, CHENG N M, YAO H, et al. Non-rigid object tracking via deformable patches using shape-preserved KCF and level sets [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 5495-5503.
- [73] BERTINETTO L, VALMADRE J, HENRIQUES J F, et

- al. Fully-convolutional siamese networks for object tracking[C]//European Conference on Computer Vision. Berlin: Springer, 2016: 850-865.
- [74] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C] // Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2012: 1097-1105.
- [75] WANG Q, GAO J, XING J, et al. Defnet: discriminant correlation filters network for visual tracking[J]. arXiv Preprint arXiv, 2017: 1704. 04057.
- [76] HE A, LUO C, TIAN X, et al. A twofold siamese network for real-time object tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 4834-4843.
- [77] GUO Q, FENG W, ZHOU C, et al. Learning dynamic siamese network for visual object tracking [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1763-1771.
- [78] HUANG C, LUCEY S, RAMANAN D. Learning policies for adaptive tracking with deep feature cascades[C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 105-114.
- [79] SONG Y, MA C, GONG L, et al. Crest: convolutional residual learning for visual tracking[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2555-2564.
- [80] ZHU Z, WU W, ZOU W, et al. End-to-end flow correlation tracking with spatial-temporal attention[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 548-557.
- [81] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 2411-2418.
- [82] WU Y, LIM J, YANG M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [83] KRISTAN M, LEONARDIS A, MATAS J, et al. The visual object tracking vot 2017 challenge results [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1949-1972.
- [84] CHOI J, JIN C H, JEONG J, et al. Visual tracking using attention-modulated disintegration and integration [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 4321-4330.
- [85] KIANI GALOOGAHI H, FAGG A, et al. Learning background-aware correlation filters for visual tracking [C] // Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 1135-1143.
- [86] GUNDOGDU E, ALATAN A. Spatial windowing for correlation filter based visual tracking [C] // 2016 IEEE International Conference on Image Processing (ICIP). Piscataway: IEEE, 2016: 1684-1688.
- [87] CHOI J, JIN C H, YUN S, et al. Attentional correlation filter network for adaptive visual tracking [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 4807-4816.

(责任编辑 吕小红)