

一种用于视觉跟踪的低秩上下文感知的相关滤波器

苏银强^{1,2} 王 宣¹ 王 淳³ 李 充³ 徐 芳¹

1 中国科学院长春光学精密机械与物理研究所 长春 130000

2 中国科学院大学 北京 100049

3 中国人民解放军空军装备部驻沈阳地区军事代表局驻长春地区军事代表室 长春 130000

(suyinqiang18@mails.ucas.ac.cn)

摘 要 基于 DCF 的目标跟踪方法在保持实时运行时,由于在精度和鲁棒性之间实现了很好的权衡而备受关注。但是,当出现遮挡、移出视野、平面外旋转等干扰时,现有跟踪器仍面临着模型漂移甚至跟踪失败的情况。为此,提出了一种基于低秩上下文感知的相关滤波器 LR_CACF。具体来说,在滤波器学习阶段,直接将目标及其上下文信息集成到 DCF 框架中,以更好地将目标从背景中鉴别出来;同时,对跨帧视频施加低秩约束以强调时序平滑性,使得学习的滤波器处于一个低维的鉴别流行上,进一步提高了跟踪性能;然后,利用 ADMM 实现滤波模型的高效优化;此外,针对模型失真的问题,启动多模态检测机制来识别响应图的可靠性,当反馈不可靠时,滤波器停止训练,同时扩大搜索区域,并采用区域重叠的方法重新捕获目标。在 OTB-50, OTB-100 和 DTB70 数据集上进行了大量实验,实验结果表明,相对于基线 SAMF_CA,在 DP 方面,LR_CACF 分别获得了 6.9%,4.0%和 7.1%的增益,AUC 分别提高了 3.6%,2.7%和 5.4%。基于属性分析的结果表明,LR_CACF 尤其擅长处理遮挡、移出视野、平面外旋转、低分辨率和快速运动等场景。

关键词: 视觉跟踪;相关滤波;低秩约束;上下文感知;重检测

中图分类号 TP394.1;TH691.9

Correlation Filter Based on Low-rank and Context-aware for Visual Tracking

SU Yinqiang^{1,2}, WANG Xuan¹, WANG Chun³, LI Chong³ and XU Fang¹

1 Changchun Institute of Optics, Fine Mechanics and Physics(CIOMP), Chinese Academy of Sciences, Changchun 130000, China

2 University of Chinese Academy of Science, Beijing 100049, China

3 The First Military Representative Office of the Military Representative Bureau of the Army Equipment Department of the Chinese People's Liberation Army in Changchun Shenyang, Changchun 130000, China

Abstract Discriminative correlation filter(DCF)-based visual tracking approaches have attracted remarkable attention due to their good tradeoff between accuracy and robustness while running at real-time. However, the existing trackers still face model drift and even tracking failure situation when there are interferences such as long-term occlusion, out-of-view and out-of-plane rotation. To this end, we propose a low-rank and context-aware correlation filter(LR_CACF). Specifically, we directly integrate the target and its global contexts into DCF framework during filter learning stage to better discriminate the target from surrounding. Meanwhile, the low-rank constraint is injected across frames to emphasize the temporal smoothness, so that the learned filter is retained in a low-dimensional discriminant manifold to further improve tracking performance. Then, the ADMM is used to optimize the model effectively. Moreover, for model distortion, the multimodal detection mechanism is utilized to identify anomaly in the response. The filter stops training while extends the search regions to recapture the target when feedback is unreliable. Finally, extensive experiments are conducted on OTB50, OTB100 and DTB70 datasets, and the results demonstrate that, compared with the baseline SAMF_CA, LR_CACF achieves gains of 6.9%, 4.0% and 7.1% in DP, respectively, and the average AUC improves by 3.6%, 2.7% and 5.4%, respectively. Meanwhile, attribute-based evaluation shows that the proposed tracker is particularly adept at handling the scenes such as occlusion, out-of-view, out-of-plane rotation, low resolution, and fast motion.

Keywords Visual tracking, Correlation filter, Low-rank, Context-aware, Redetection

到稿日期:2023-07-07 返修日期:2023-10-10

基金项目:国家自然科学基金面上项目(62175233);吉林省自然科学基金面上项目(2022010111JC)

This work was supported by the General Program of National Natural Science Foundation of China(62175233) and General Program of Natural Science Foundation of Jilin Province, China(2022010111JC).

通信作者:王宣(lly637@163.com)

1 引言

视觉目标跟踪是计算机视觉领域最基本的研究课题之一,已被广泛应用于智能视频监控和军事打击等领域。其主要任务是在视频或者图像序列中,给定任意对象的初始状态(如尺度和中心位置),在后续序列中持续、稳定、精确地估计目标运动轨迹和状态^[1-3]。经过几十年的发展,大量的目标跟踪方法相继被提出,并在近年来取得了显著的成果。但是,在实际跟踪过程中,目标可能被其他物体部分/完全遮挡,或者尺度发生变化,以及受背景杂波、光照变化、运动模糊等干扰的影响,很容易出现模型漂移甚至跟踪失败的情况。因此如何在无约束条件下设计一种鲁棒、精确的目标跟踪算法仍然是一个具有挑战性的研究课题。

现有的视觉目标跟踪方法主要包括鉴别式相关滤波(DCF)和孪生网络(SN)两类框架^[4-5]。基于 SN 的方法通过最大化目标外观与背景之间的距离,同时最小化两个背景块和目标本身的距离,来离线学习一个嵌入空间^[6-7]。这类方法虽然有效,但复杂的卷积运算限制了其在实际应用中的发展。相反,基于 DCF 的方法在保持实时运行的同时,在精度和鲁棒性之间实现了很好的权衡,因而备受关注。这类方法的核心思想是将跟踪问题转换成具有循环结构的岭回归问题,通过最小化最小二乘损失,同时利用大量的循环移位样本在线学习滤波器,然后将学习得到的滤波器应用于感兴趣区域,最后通过最大响应来估计目标状态。这种范式的成功主要归结于两方面:1)通过循环移位基础样本来近似实现密集采样,弥补了滤波器训练过程中样本不足的问题;2)利用循环矩阵在频域对角化的性质,规避了复杂的矩阵求逆运算,从而降低了计算复杂度,提高了跟踪速度。

基于 DCF 的方法利用循环移位和频域快速计算技术实现了高效跟踪。为了进一步提升跟踪性能,研究者们主要从 3 个方面进行改进:1)融合不同特征表征目标^[8-9],如 HOG, CN 等人工设计的特征以及 CNN 提取的卷积特征;2)加入尺度估计策略^[10-11],有效处理目标尺度变化;3)改进外观模型^[12-13],在经典的 DCF 框架中引入空间或时间正则约束,进一步提升跟踪性能。但是,训练样本的周期性假设,使得边界处的样本不连续,从而产生不必要的空间边界效应,当快速运动,存在背景杂波、遮挡等干扰时,很容易出现模型漂移甚至跟踪失败的情况。此外,大多数滤波器在构建目标模型时只考虑当前帧信息,没有充分利用历史信息,目标尺度变化、非刚性形变、旋转等因素会导致目标外观发生变化,进而很容易导致模型退化,降低跟踪性能。

为了有效处理上述问题,本文提出了一种基于低秩上下文感知的目标跟踪算法。具体来说,在构建目标外观模型时,同时考虑了目标的时间和空间信息。首先,目标及其周围的上下文信息被集成到 DCF 框架中,以缓解边界效应,提高模型在遇到遮挡、相似目标等干扰下的鉴别力;此外,对跨帧视频施加了低秩约束,使得学习到的滤波器处于低维的鉴别流行上面,以获得更加稳定、鲁棒的滤波模型;然后,针对模型失真,对响应图进行多模态检测,检测响应图是否存在异常,根据反馈的检测结果,来指导滤波器的高置信度更新。当响应图存在多峰时,滤波器停止更新,同时在目标周围扩大搜索区域以重新捕获目标,进一步缓解模型漂移甚至跟踪失败的情况。

最后,在 OTB-50, OTB-100 和 DTB70 数据集上进行大量实验,评估结果证明了所提方法的有效性。

LR_CACF 跟踪器的整体框架如图 1 所示。正常情况下,通过响应峰值确定目标状态;当异常出现时,基于区域重叠的方法在目标周围 4 个方向(上、下、左、右)重采样以重新捕获目标。此外,通过与历史峰值均值的比较,并采用 Kalman 滤波,进一步避免模型漂移。最后,通过多峰检测机制实现跟踪器和重检测器的高度协调。

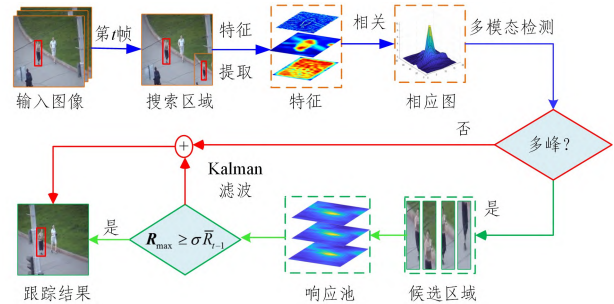


图 1 LR_CACF 跟踪器的框架图

Fig. 1 Framework of LR_CACF tracker

本文的创新点可以总结为如下 3 个方面:

- 1) 构建了一种基于低秩上下文感知的相关滤波模型,有效提高了模型在相似目标干扰、背景杂波等复杂场景下的鉴别能力。
- 2) 利用 ADMM 技术推导出 LR_CACF 模型的闭式解,在不牺牲跟踪效率的前提下获得了较好的跟踪性能。
- 3) 提出了一种基于区域重叠的重捕获机制,能够很好地应对遮挡、移出视野等场景。

2 研究现状

近年来,基于 DCF 的目标跟踪方法因在多个跟踪基准上取得了显著成果而备受关注。MOSSE^[14]首次将信号处理领域的相关滤波引入目标跟踪中,采用较少的样本以及灰度特征实现了自适应跟踪。CSK^[15]通过循环移位基础样本来近似密集采样,同时利用训练样本的循环特性,在频域中进行高效计算,但是,其仍然采用灰度特征。随后,KCF^[8]引入高斯核函数和 HOG 特征,高斯核函数将非线性问题转换为高维空间中的线性可分问题;HOG 特征有效地表征目标的纹理和轮廓信息。同时,fDSST^[11]采用多通道的 CN 特征来替代灰度特征,此外,为了提高计算效率,采用 PCA 方法对特征进行降维,实现了实时跟踪。SAMF^[10]将灰度、HOG、CN 特征进行矢量链接来表征目标。Staple 利用 HOG 和颜色统计特征的互补特性,对两种特征进行自适应融合,实现稳定跟踪。近年来,学者们将 CNN 特征应用于目标跟踪领域。HCFT^[9]通过充分分析不同卷积特征,利用 VGGNet19 网络提取出来的 Conv3-4, Conv4-4, Conv5-4 层的输出表征目标。实际上,这类网络是专门为图像分割任务而设计的,直接将 CNNs 提取的卷积特征应用于目标跟踪仍然具有挑战性。

在实际跟踪中,平台高度或相机焦距的变化以及目标形变等都有可能造成目标尺度发生变化。为了应对这些问题,RPCA^[16]在目标周围选择与目标等大的 4 个上下文块,通过判断 4 个块的中心位置与目标中心的距离来估计目标的尺度。SAMF^[10]采用穷尽尺度搜索策略,在 7 个固定缩放因子

下对目标进行采样,然后利用最大响应同时估计目标尺度和位置。与 SAMF 不同,fDSST^[11]在位置滤波器之外单独训练一个尺度滤波器,有效地估计目标尺度变化。由于这种尺度估计策略独立于位置滤波器之外,因此可以有效地移植到其他没有尺度估计模块的跟踪器中。

基于 DCF 的跟踪方法在融合不同特征、引入高斯核函数和加入尺度估计模块之后,取得了显著的成果。但是由于训练样本周期性假设,产生了不必要的空间边界效应,降低了滤波模型的鉴别性。为解决上述问题,SRDCF^[13]引入了空间正则约束,具体地,根据滤波器系数的空间位置进行惩罚,距离滤波器中心较近的位置约束越小,反之亦然。随后,Danelljan 等又提出了 C-COT^[17]和 ECO^[18],进一步缓解了边界效应。BACF^[12]通过扩大搜索区域,引入二值裁剪矩阵,在增加样本数量的同时提高了样本的质量。CACF^[19]将目标周围的上下文信息集成到 DCF 框架中,有效抑制了快速运动、遮挡和背景杂波等复杂场景下的模型漂移问题。为了进一步提升跟踪性能,Li 等通过引入时间正则项融合了历史信息,进一步优化了 SRDCF,有效缓解了模型退化^[20]的问题。随后,AutoTrack^[21]通过充分分析响应图的局部-全局信息,提出了一种自适应时空正则滤波器。Xu 等^[3]利用自适应属性感知空间注意机制构建外观模型。Jain 等^[22]提出了一种通道注意模型,并研究了两种不同的注意正则化方法,以强调重要的特征通道。AS2RCF 跟踪器^[23]引入了 l_1 范数,有效地抑制了混合在模型中的噪声,使得滤波器在保持稳定的同时具有一定的稀疏性。

相比经典的 DCF 框架,加入正则化的滤波器能够有效地改善跟踪性能,但是,性能的改进都是以牺牲跟踪速度为代价的,现在的 DCF 滤波器逐渐不具有实时跟踪性能。如何在保持实时运行的同时进一步提高跟踪器的鲁棒性和精度仍然具有重要的研究价值。

3 本文方法

3.1 核化相关滤波

在详细阐述本文提出的滤波模型之前,为了文章的完整性,首先回顾经典的核相关滤波 KCF。KCF 将跟踪问题解释为二分类问题,直接从复杂的背景中将目标鉴别出来。其可以公式化成如下目标函数:

$$\arg \min_{\mathbf{w}} \left\| \sum_{i=1}^n \mathbf{w}^T \mathbf{x}_i - \mathbf{y} \right\|_2^2 + \lambda_1 \left\| \mathbf{w} \right\|_2^2 \quad (1)$$

其中, $\mathbf{w} \in \mathbb{R}^n$ 表示需要优化的滤波器, $\mathbf{x}_i \in \mathbb{R}^n (i=1,2,\dots,n)$ 表示向量化的训练样本,正则项参数 λ_1 用于控制滤波器过拟合的程度。式(1)岭回归项中 $\sum_{i=1}^n \mathbf{w}^T \mathbf{x}_i$ 可以进一步表示为 $\mathbf{X}\mathbf{w}$, 其中 $\mathbf{X}$ 表示数据矩阵,其每一行表示一个训练样本,即:

$$\mathbf{X} = C(\mathbf{x}) = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_1[\Delta\tau_1] \\ \vdots \\ \mathbf{x}_1[\Delta\tau_{n-1}] \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{12} & x_{13} & \cdots & x_{1n-1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{11} & \cdots & x_{1n-1} \end{pmatrix}$$

目标函数可进一步表示为:

$$\arg \min_{\mathbf{w}} \left\| \mathbf{X}\mathbf{w} - \mathbf{y} \right\|_2^2 + \lambda_1 \left\| \mathbf{w} \right\|_2^2 \quad (2)$$

为了优化目标函数,对式(2)进行求导,并令导数等于 0,得到如下闭式解:

$$\mathbf{w} = (\mathbf{X}^H \mathbf{X} + \lambda_1 \mathbf{I})^{-1} \mathbf{X}^H \mathbf{y} \quad (3)$$

其中, $\mathbf{X}^H$ 表示矩阵 $\mathbf{X}$ 的 Hermitian 转置, $(a)^{-1}$ 表示求逆运算。

由于式(3)中 $\mathbf{X}^H$ 和 $\mathbf{X}$ 均为循环矩阵,因此可以利用循环矩阵在频域对角化的性质:

$$\mathbf{X} = \mathbf{F} \text{diag}(\hat{\mathbf{x}}) \mathbf{F}^H \quad (4)$$

$$\mathbf{X}^H = \mathbf{F} \text{diag}(\hat{\mathbf{x}}^*) \mathbf{F}^H$$

利用式(4)可以得到如下等式:

$$\mathbf{X}^H \mathbf{X} = \mathbf{F} \text{diag}(\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}}) \mathbf{F}^H \quad (5)$$

联合式(5)和式(3),得到滤波器在频域的最优解:

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}^* \odot \hat{\mathbf{y}}}{\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}} + \lambda} \quad (6)$$

其中,分数表示矩阵的点除运算,符号 $\odot$ 表示矩阵的 Hadamard 积。式(6)中只涉及矩阵元素的点乘和点除操作,巧妙地规避了矩阵的逆运算,有效提高了运行速度。

这里需要特别说明的是,在本文中,为了方便表示,采用小写字母表示常量,粗体小写字母表示向量,粗体大写字母表示矩阵。

3.2 低秩上下文感知的滤波模型

在跟踪过程中,如果目标周围出现与目标纹理和颜色信息相似的物体,或被其他物体遮挡时,目标外观很容易发生退化。并且,如果伴随旋转、尺度变化等,这些干扰的叠加会使得目标外观发生退化,最终可能导致模型漂移甚至跟踪失败。基于这种观察,在构建滤波模型时,我们充分利用了目标周围的背景信息,在滤波器学习阶段直接将上下文信息集成到 DCF 框架中;此外,为了适应目标和周围背景的动态变化,对相邻两帧的滤波器施加低秩约束,使得学习的滤波器处于一个低维的鉴别流行上面,以获得更加稳定的模型,目标函数如下:

$$\arg \min_{\mathbf{w}} \left\| \mathbf{X}\mathbf{w} - \mathbf{y} \right\|_2^2 + \lambda_1 \left\| \mathbf{w} \right\|_2^2 + \lambda_2 \sum_{i=1}^k \left\| \mathbf{X}_i \mathbf{w} \right\|_2^2 \quad (7)$$

$$\text{s. t. } \text{rank}(\mathbf{W}) - \text{rank}(\mathbf{W}_{t-1}) < \eta$$

其中,矩阵 $\mathbf{X}_i \in \mathbb{R}^{n \times n} (i \in [1, k])$ 表示目标周围的上下文块张成的数据矩阵,常数 $\lambda_2 > 0$ 用于控制背景响应回归到零的正则项参数, $\text{rank}()$ 表示矩阵的求逆运算。 $\mathbf{W} = \left[\frac{t}{t-1} (1-\alpha) \mathbf{W}_{t-1}, \text{taw} \right] (t > 1)$ 存储了历史帧的滤波器,常数 α 表示滤波器的学习率。式(7)中涉及矩阵秩的优化,虽然矩阵秩的优化是一个 NP 难问题,但幸运的是,跨视频帧的低秩约束可以用它的充分条件代替。目标函数式(7)的优化等价于:

$$\arg \min_{\mathbf{w}} \left\| \mathbf{X}\mathbf{w} - \mathbf{y} \right\|_2^2 + \lambda_1 \left\| \mathbf{w} \right\|_2^2 + \lambda_2 \sum_{i=1}^k \left\| \mathbf{X}_i \mathbf{w} \right\|_2^2 + \mu \left\| \mathbf{w} - \bar{\mathbf{w}} \right\|_2^2 \quad (8)$$

其中, $\bar{\mathbf{w}}$ 表示滤波器滑动平均更新的结果,即 $\bar{\mathbf{w}} = \alpha \mathbf{w} + (1-\alpha) \bar{\mathbf{w}}_{t-1}$,目标函数式(8)中只需要存储一帧的滤波器,减轻了存储压力。

由于式(8)是一个凸函数,因此可以采用增广 Lagrange 乘子法进行优化。首先引入辅助变量 $\mathbf{w} = \mathbf{g}$,目标函数式(8)表示为含有等式约束的优化问题:

$$\arg \min_{\mathbf{w}} \left\| \mathbf{X}\mathbf{w} - \mathbf{y} \right\|_2^2 + \lambda_2 \sum_{i=1}^k \left\| \mathbf{X}_i \mathbf{w} \right\|_2^2 + \lambda_1 \left\| \mathbf{w} \right\|_2^2 + \mu \left\| \mathbf{w} - \bar{\mathbf{w}}_{t-1} \right\|_2^2 \quad (9)$$

s. t. $\mathbf{w} = \mathbf{g}$

构造式(9)的增广 Lagrangian 函数如下:

$$\arg \min_{\mathbf{w}} \left\| \mathbf{X}\mathbf{w} - \mathbf{y} \right\|_2^2 + \lambda_2 \sum_{i=1}^k \left\| \mathbf{X}_i \mathbf{w} \right\|_2^2 + \lambda_1 \left\| \mathbf{g} \right\|_2^2 +$$

$$\mu \left\| \mathbf{w} - \bar{\mathbf{w}}_{t-1} \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{w} - \mathbf{g} + \frac{\mathbf{s}}{\gamma} \right\|_2^2 \quad (10)$$

其中, $\mathbf{s}$ 表示 Lagrangian 乘子, γ 表示惩罚参数。为了保证解的收敛性, 采用 ADMM 技术将目标函数等效分解成如下 3 个子问题:

$$\begin{cases} \arg \min_{\mathbf{w}} \left\| \mathbf{X}\mathbf{w} - \mathbf{y} \right\|_2^2 + \lambda_2 \sum_{i=1}^k \left\| \mathbf{X}_i \mathbf{w} \right\|_2^2 + \\ \mu \left\| \mathbf{w} - \bar{\mathbf{w}}_{t-1} \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{w} - \mathbf{g} + \frac{\mathbf{s}}{\gamma} \right\|_2^2 \\ \arg \min_{\mathbf{g}} \lambda_1 \left\| \mathbf{g} \right\|_2^2 + \frac{\gamma}{2} \left\| \mathbf{w} - \mathbf{g} + \frac{\mathbf{s}}{\gamma} \right\|_2^2 \\ \mathbf{s} = \mathbf{s} + \gamma(\mathbf{w} - \mathbf{g}) \end{cases} \quad (11)$$

优化与求解子问题 $\mathbf{w}$: 给定 $\mathbf{g}$ 和 $\mathbf{s}$, 对子问题 $\mathbf{w}$ 进行微分并令导数等于 0, 得到下式:

$$\mathbf{w} = \left(\mathbf{X}_0^H \mathbf{X}_0 + \lambda_2 \sum_{i=1}^k \mathbf{X}_i^H \mathbf{X}_i + \mu + \frac{\gamma}{2} \right)^{-1} \left(\mathbf{X}_0^T \mathbf{y} + \mu \bar{\mathbf{w}}_{t-1} + \frac{\gamma}{2} \mathbf{g} - \frac{\mathbf{s}}{2} \right) \quad (12)$$

其中, 数据矩阵 $\mathbf{X}_0$ 和 $\mathbf{X}_i$ ($i \in [1, k]$) 均为循环矩阵, 利用式(12)循环矩阵可以在频域对角化, 得到如下等式:

$$\mathbf{X}_0^H \mathbf{X}_0 = \mathbf{F} \text{diag}(\hat{\mathbf{x}}_0^* \odot \hat{\mathbf{x}}_0) \mathbf{F}^H \quad (13)$$

$$\mathbf{X}_i^H \mathbf{X}_i = \mathbf{F} \text{diag}(\hat{\mathbf{x}}_i^* \odot \hat{\mathbf{x}}_i) \mathbf{F}^H$$

联合式(12)和式(13), 得到如下等式:

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{x}}_0^* \odot \hat{\mathbf{y}} + \mu \hat{\mathbf{w}}_{t-1} + \frac{\gamma}{2} \hat{\mathbf{g}} - \frac{\hat{\mathbf{s}}}{2}}{\hat{\mathbf{x}}_0^* \odot \hat{\mathbf{x}}_0 + \lambda_2 \sum_{i=1}^k \hat{\mathbf{x}}_i^* \odot \hat{\mathbf{x}}_i + \mu + \frac{\gamma}{2}} \quad (14)$$

优化与求解子问题 $\mathbf{g}$: 给定 $\mathbf{w}$ 和 $\mathbf{s}$, 优化 $\mathbf{g}$, 对子问题 $\mathbf{g}$ 进行求导, 得到如下等式:

$$\mathbf{g} = \frac{\gamma \mathbf{w} + \mathbf{s}}{2\lambda_1 + \gamma} \quad (15)$$

优化与求解子问题 $\mathbf{s}$: 给定 $\mathbf{w}$ 和 $\mathbf{g}$, 优化 $\mathbf{s}$, 更新 Lagrangian 乘子 $\mathbf{s}$ 和惩罚参数 γ :

$$\begin{aligned} \mathbf{s} &= \mathbf{s} + \gamma(\mathbf{w} - \mathbf{g}) \\ \gamma &= \min(\beta\gamma, \gamma_{\max}) \end{aligned} \quad (16)$$

其中, 迭代步长 β 用于控制 $\mathbf{w}$ 和 $\mathbf{g}$ 的近似程度, 惩罚参数的最大值 $\gamma_{\max}$ 用于防止滤波器最优解奇异。

3.3 后处理机制

基于 DCF 的目标跟踪方法通过在线学习一个相关滤波器, 并将学习到的滤波器应用于下一帧的感兴趣区域, 然后根据响应图最大值的位置来定位目标的状态, 这种假设在很多情况下是满足的。但是, 当目标外观持续变化或目标周围存在相似的颜色、纹理时, 响应图可能会出现意外峰值, 如果不考虑响应的可靠性, 直接采用峰值估计目标状态, 很容易产生灾难性后果。为此, 在 LR_CACF 模型之外, 加入多模态检测机制, 如果响应图中存在多个峰值时, 跟踪器将扩大搜索范围并重新搜索目标。

具体地, 当获取当前帧的响应图 $\mathbf{R}_t$ 之后, 将响应图的零频分量移动到数据中心, 重新排列响应图 $\mathbf{R}_t$, 公式化为:

LR_CACF 跟踪算法伪代码如算法 1 所示。

$$\mathbf{R}_{i,j} = \mathbf{R}_{i,j} \left[\nabla \tau \left(\text{round} \left(\frac{h}{2} \right) \right) \right] \left[\nabla \tau \left(\text{round} \left(\frac{w}{2} \right) \right) \right] \quad (17)$$

其中, $\nabla \tau[k]$ 表示循环移位操作, w 和 h 分别表示响应图的宽

度和高度。计算响应值的峰值 $peak$ 和次峰值 $peak_s$, 峰值与次峰值关系满足等式 $peak_s = \tau peak$, 其中 $\tau \in [0, 1]$ 为常数。然后筛选峰值 $peak$ 和次峰值 $peak_s$ 之间的响应值, 形成一个列表, 列表中存储响应值 r 及对应的位置 (x_r, y_r) 。同时, 将小于次峰值的响应全部设置为 0, 得到新的响应图 $\mathbf{R}_N$; 最后, 在新的响应图中遍历列表, 来确定响应图中是否存在多个峰值, 具体计算式如下:

$$P(R_r) = \mathbf{R}(x_r, y_r) \odot \mathbf{B} \quad (18)$$

其中, 矩阵 $\mathbf{B} \in R^{m \times m}$ 是一个二值矩阵, 用于确定局部最大值; $\mathbf{R}(x_r, y_r)$ 表示在响应图中以 (x_r, y_r) 为中心, 选取与矩阵 $\mathbf{B} \in R^{m \times m}$ 等大的局部响应图; $P(R_r)$ 表示当前位置是否为多峰。如果响应图呈现多峰状态, 跟踪器将在目标周围扩大搜索区域, 并且采用 Kalman 滤波辅助的方法, 重新捕获目标。

首先, 以上一帧估计的目标中心为基准, 上下左右进行多次相互独立的采样, 其中每次采样的区域具有一定的重合度。如果采样区域的最大响应 $\mathbf{R}_{\max}$ 满足如下关系:

$$\mathbf{R}_{\max} \geq \sigma \bar{\mathbf{R}}_{t-1} \quad (19)$$

则选取最大响应区域定位目标, 如果式(19)不成立, 则采用 Kalman 滤波跟踪目标。

算法 1 LR-CACF 跟踪算法

输入: 当前帧图像 I_t , 第 $t-1$ 帧目标状态 $(x_{t-1}, y_{t-1}, w_{t-1}, h_{t-1})$, 滤波器 $\bar{\mathbf{w}}_{t-1}$

输出: 第 t 帧目标状态 (x_t, y_t, w_t, h_t) , 滤波器 $\bar{\mathbf{w}}_t, \bar{\mathbf{w}}_t^{D-G}$

1. Repeat
2. 以 (x_{t-1}, y_{t-1}) 为中心裁剪多尺度图像, 并提取其特征表示 $\{\mathbf{z}_i^k\}_{k=1}^7$;
3. 通过 $\mathcal{F}^{-1}(\hat{\mathbf{z}}_i^k \odot \hat{\mathbf{w}}_{t-1})$ 计算多尺度响应 $\{\mathbf{R}_k\}_{k=1}^7$;
4. 通过优化目标函数 $\arg \min(\mathbf{R}_1, \dots, \mathbf{R}_7)$ 计算最大响应 $\mathcal{R}_{\max}$ 值对应的层 $\mathcal{R}_m$;
5. 对响应图 $\mathbf{R}_m$ 进行多峰检测;
6. If 响应图 $\mathbf{R}_m$ 存在多峰值
7. 根据目标状态 $(x_{t-1}, y_{t-1}, w_{t-1}, h_{t-1})$ 扩大搜索区域, 同时滤波器停止更新;
8. 提取每个区域的特征 $\{\mathbf{z}_i\}_{i=1}^5$, 并计算其响应 $\{\mathbf{R}_i\}_{i=1}^5$;
9. If 满足条件: $\max \mathbf{R}^i \geq \delta \bar{\mathbf{R}}_{t-1}$
10. 根据确定 $\max \mathbf{R}^i$ 目标中心 (x_t, y_t) , 尺度不变 $(w_t, h_t) = (w_{t-1}, h_{t-1})$;
11. Else
12. Kalman 滤波估计目标位置 (x_t, y_t) , 尺度不变 $(w_t, h_t) = (w_{t-1}, h_{t-1})$;
13. Endif
14. Else
15. 根据最大响应 $\mathbf{R}_m$ 确定目标状态 (x_t, y_t, w_t, h_t) ;
16. 采用 ADMM 方法, 基于式(12)~式(16)迭代优化滤波器 w_t
17. 更新滤波器: $\bar{\mathbf{w}}_t = (1-\eta)\bar{\mathbf{w}}_{t-1} + \eta w_t$
18. End
19. 最后一帧

4 实验部分

为验证提出的 LR_CACF 的有效性, 将其在 OTB50, OTB100 和 DTB70 数据集上与其他最先进的跟踪器进行比较。下面将首先详细介绍实现细节和参数设置。然后在不同的数据集上进行量化和质化实验, 并进行全面的分析和讨论,

以验证跟踪器的优势及不足。

4.1 实验细节和参数设置

LR_CACF 采用 HOG, CN 和 IC 等人工特征, 以及 VGG-Net-19 网络 Pool-1 和 Pool-2 层的输出作为深度特征来表示目标; HOG 特征采用 9 个方向、 4×4 单元大小的网格, CN 和 IC 也使用 4×4 单元大小来统一单元大小。对于深度特征, 将给定搜索窗口 $M \times N$ (目标尺度的 2.5 倍) 大小的图像帧输入 VGGNet-19 网络中, 提取 Pool-1 和 Pool-2 输出, 并将输出的大小重新调整为 $M/4 \times N/4$ 。

本文涉及的实验参数如下: 正则项参数 $\lambda_1 = 10$, $\mu = 16$, 对于控制背景回归到零的正则项参数, 人工特征和深度特征对应参数分别设置为 $\lambda_2 = 10^{-4}$ 和 $\lambda_2 = 5$ 。对于学习率 α , 人工特征和深度特征分别设置为 0.005 和 0.01。需要特别说明的是, 由于训练 VGGNet-19 网络所用的 ImageNet 数据集全部为彩色图像, 因此, 当图像为灰度图像时, 只采用人工特征表示目标。

本文提出的算法的所有实验都是基于 Intel(R) Core (TM) i7-10700F CPU @ 2.90 GHz 2.90 GHz 处理器, 16 GB 的 RAM 的 64 位操作系统的 PC 机上实现。

4.2 测试数据集和评价指标

本文提出的跟踪器在 OTB50, OTB100 和 DTB70 数据集上进行测试。其中 DTB70 数据集由 70 个航拍视频序列组成, 专注于通过设计相机运动跟踪地面车辆和人。采用一次通过评估 (One-Pass Evaluation OPE) 方法, 使用精度跟踪

曲线和成功跟踪曲线评估 LR_CACF 的有效性。用于对比的跟踪器包括 CSK, fDSST, SRDCF, BACF, ROT, SAMF_CA, Staple 和 STAPLE_CA。

4.3 量化实验结果与分析

4.3.1 整体性能分析

图 2 展示了 LR_CACF 与其他 9 个最先进的跟踪器在 OTB50 数据集上的跟踪精度曲线和重叠成功曲线。总体来说, LR_CACF 在精度方面优于其他跟踪器, 但是 AUC 得分比排名第一的 ARCF 略低 0.6%。从数值结果上看, LR_CACF 获得了 79.1% 的 DP 和 54.9% 的 AUC 得分。相比 STAPLE_CA 和 SAMF_CA 这两种上下文感知的滤波器, 在 DP 指标方面, LR_CACF 分别获得了 4.8% 和 6.9% 的增益; 在 AUC 方面, 分别获得了 0.7% 和 3.9% 的增益。从图 2 的跟踪成功曲线可以看出, 当阈值较低时, LR_CACF 的 OP 得分优于其他跟踪器, 随着阈值的增加, OP 值逐渐降低, 最终导致 AUC 得分略低于排名第一的 APCF 跟踪器。具体原因有两个: 一是对背景约束的强假设, 二是尺度池的策略使得搜索范围减小。但是, 实验结果表明, LR_CACF 预测的目标中心没有偏离实际的目标中心。

图 3 展示了 LR_CACF 在 OTB100 数据集上的跟踪成功曲线和成功重叠曲线。实现结果表明, 本文提出的方法几乎取得了最好的性能, 精度比排名第二的 STAPLE_CA 跟踪器高出 2.3%, 相比基准 SAMF_CADP 和 AUC 得分分别高出 4% 和 2.6%。

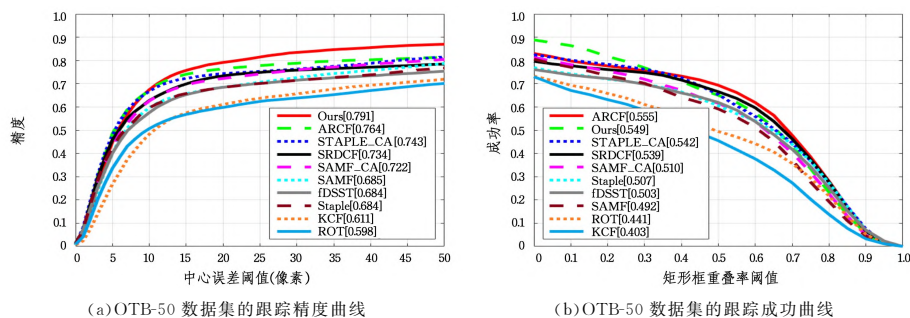


图 2 LR_CACF 跟踪器在 OTB-50 上的精度和成功曲线

Fig. 2 Precision and success plots of LR_CACF tracker on OTB-50

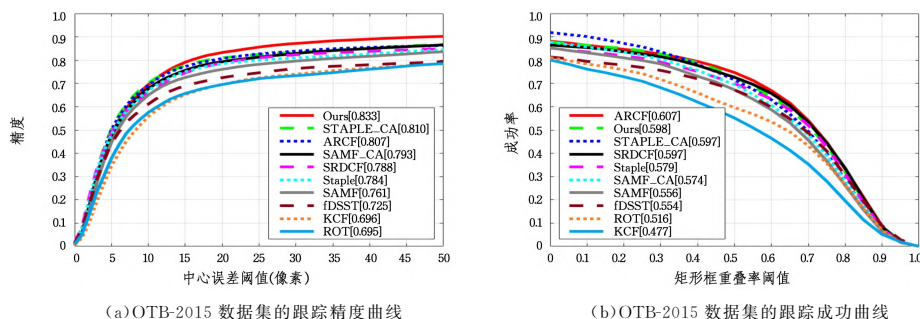


图 3 LR_CACF 跟踪器在 OTB100 上的精度和成功曲线

Fig. 3 Precision and success plots of LR_CACF tracker on OTB100

图 4 展示了 LR_CACF 跟踪器在 DTB70 数据集上的评估结果。可以看出 LR_CACF 无论是在精度还是成功率方面都获得了最佳的性能, 优于其他经典的 DCF 跟踪器, 平均精度和成功率分别为 60.3% 和 39.9%, DP 和 AUC 得分分别领先于排名第二的 BACF 算法 4.1% 和 1.1%。相比基线跟踪

SAMF-CA, 所提算法在精度和成功率方面分别获得了 7.1% 和 5.4% 的增益。SRDCF 的 DP 值为 51.2%, AUC 得分为 36.3%, 比 LR_CACF 算法低 9.1% 和 3.6%。与基于颜色直方图和 HOG 特征互补的跟踪器 Staple 相比, LR_CACF 性能更好。

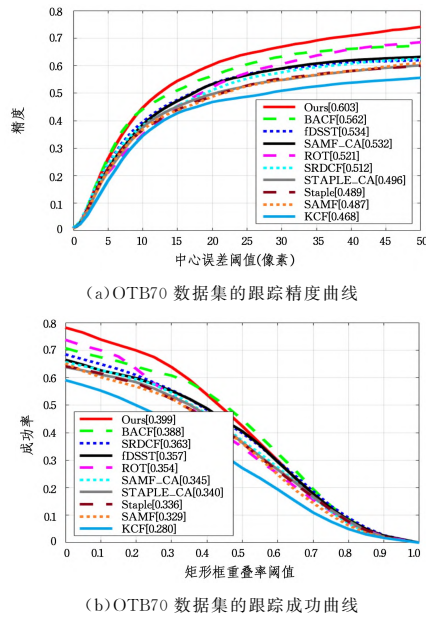


图 4 LR_CACF 跟踪器在 DTB70 上的精度和成功曲线

Fig. 4 Precision and success plots of LR_CACF tracker on DTB70

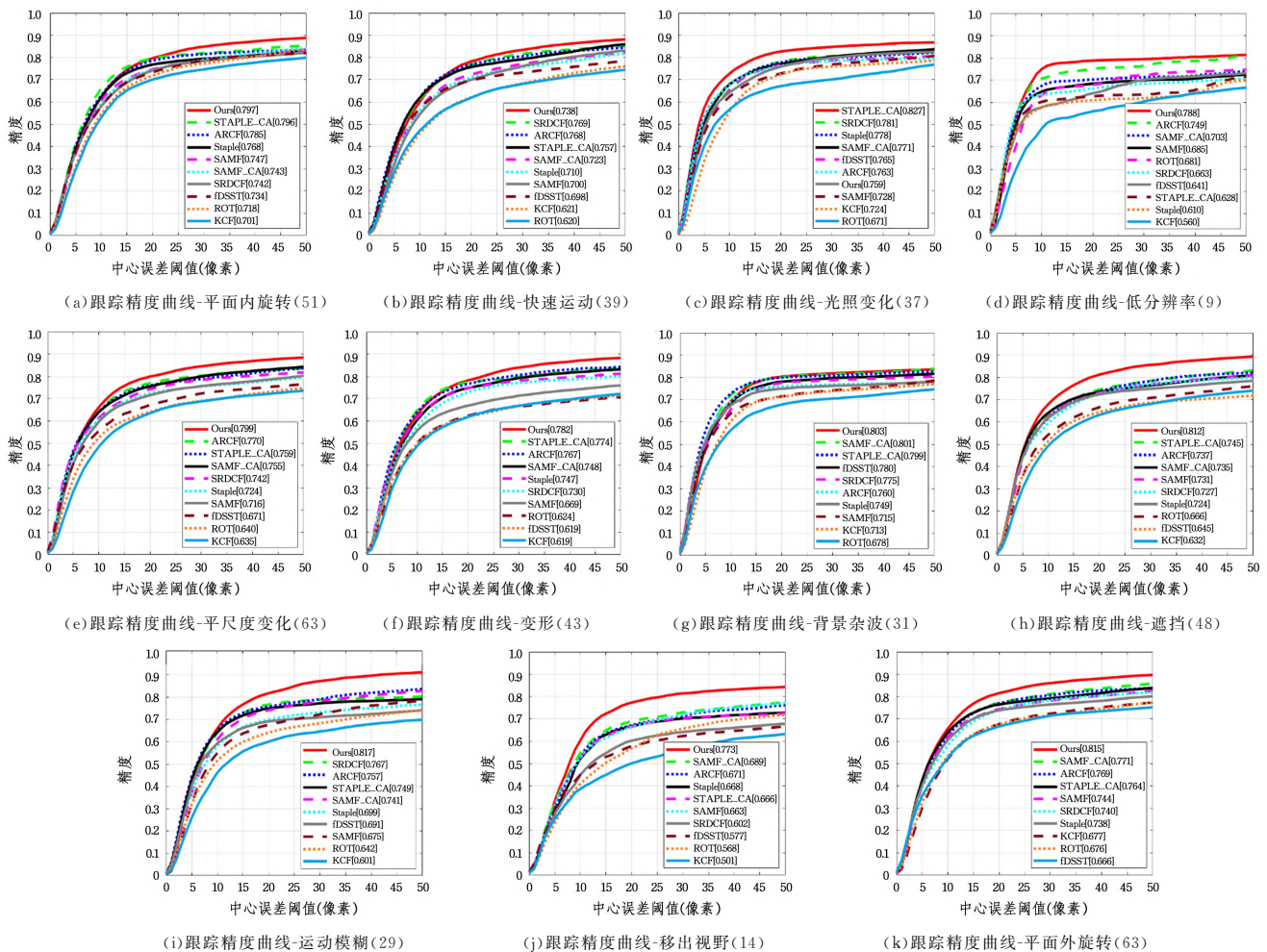


图 5 LR_CACF 跟踪器在 OTB100 的 11 种属性上的精度曲线

Fig. 5 Precision curves of LR_CACF tracker on 11 attributes of OTB100

实验结果表明,除光照变化之外,LR_CACF 跟踪器在其他属性上均取得了最好的性能。其中,在 OPR 和 IPR 上,LR_CACF 跟踪器分别获取了 81.5% 和 79.7% 的 DP 得分,并且,

总体而言,在 OTB50, OTB100 和 DTB70 这 3 个数据集上的实验结果表明,本文提出的 LR_CACF 无论是在精度还是在成功率方面都获得了较好的性能。但是,从表 1 可以看出,LR_CACF 跟踪器在 OTB100 数据集上的速度大约在 1.79 FPS 左右,无法满足实时需求。主要原因是采用 VGG-Net 网络提取的 CNN 特征表示目标网络层数较深,运算比较耗时。下一步将在低秩上下文感知模型的基础上,尝试利用轻量化 CNNs 网络提取特征,进一步优化 LR_CACF。

表 1 LR_CACF 在 3 个数据集上的跟踪性能

Table 1 Tracking performance of LR_CACF on 3 datasets

Dataset	OP/%	AUC/%	DP/%	FPS/s
OTB50	65.4	54.9	79.1	1.80
OTB100	71.5	59.8	83.3	1.79
DTB70	42.4	39.9	60.3	1.45

4.3.2 基于属性的分析

为了进一步分析 LR_CACF 性能,对其进行基于属性的分析。图 5 显示了 LR_CACF 和其他 9 个跟踪器在 OTB100 数据集的 11 种不同属性上的精度曲线。

在 DEF 和 MF 挑战因素中,LR_CACF 跟踪器比基准跟踪器分别获得了 4.3% 和 7.6% 的增益,这得益于其在构建目标模型时对相邻两帧视频之间施加低秩平滑约束,使得学习的

外观模型处在一个低维的鉴别流行上面,当目标外观变化较大时,LR_CACF 提供了一个稳定的外观模型。

在 SV 属性上,LR_CACF 比专注于尺度估计的 fDSST 高出 12.6% 的 DP。SAMF_CA、STAPLE_CA 和本文的跟踪器都对背景进行了约束,在 BC 上,这 3 个跟踪器均取得了良好的性能。在面对 OCC 和 OV 挑战时,后处理机制有效地避免了模型漂移和跟踪失败的情况,LR_CACF 比具有重检测模块的 ROT 的 DP 高 14.6% 和 20.5%,比基线跟踪器 SAMF_CA 高 5.2% 和 8.4%。此外,在 9 个低分辨率图像上 LR_CACF 获得了 78.8% 的 DP,这些低分辨率图像主要以灰度图像为主,目标特征较少,随机噪声较多。

但是,上下文感知的低秩模型也很好地避免了模型漂移。虽然 LR_CACF 模型在很多挑战上表现良好,但面对光照时效果欠佳,主要原因是跟踪器施加了低秩约束,使其不能很好地适应外界环境的变化。

4.4 质化实验结果与分析

图 6 展示了 LR_CACF 与 SAMF_CA、BACF、STAPLE_CA 和 ROT 跟踪器在 OTB100 数据集的 8 个视频序列上的视觉效果。

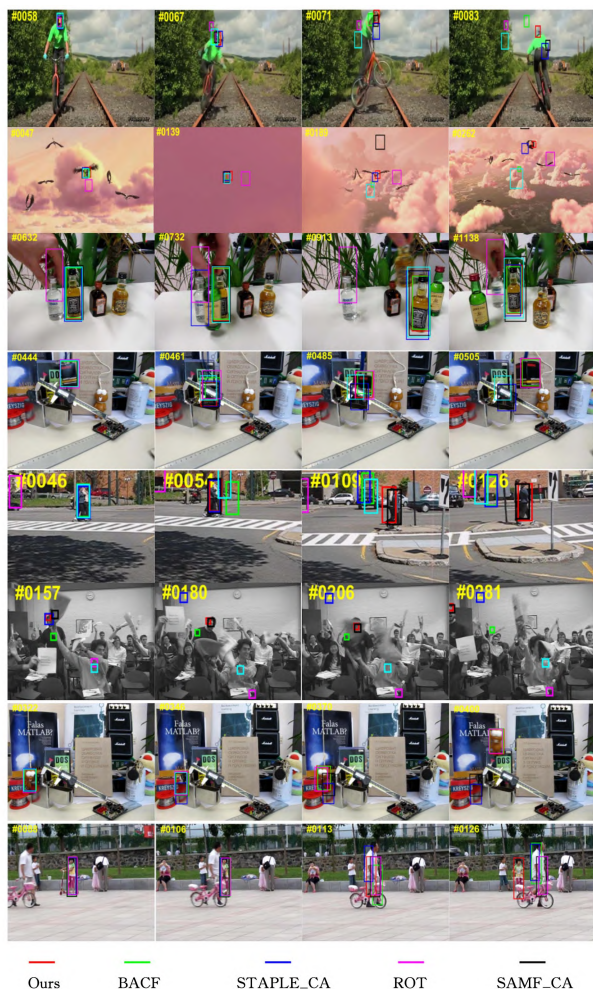


图 6 LR_CACF 与 4 个最先进的跟踪器的视觉对比

Fig. 6 Visual comparison of LR_CACF with 4 state-of-the-art trackers

这 8 个序列从上到下依次为 Biker, Bird1, Liquor, Box,

couple, Human3, Lemming 和 Girl2。在 Girl2 序列中,目标 'girl2' 被其他物体长时间遮挡(大约 16 帧),其他跟踪器如 SAMF_CA、BACF、STAPLE_CA 和 ROT 均丢失了目标,而 LR_CACF 一直稳定跟踪着目标,自始至终没有造成模型漂移,这得益于多峰检测机制和重搜索策略。序列 Box 和 Lemming 上也出现了同样的问题,但 LR_CACF 也没有造成跟踪失败的情况。此外,序列 'Bird1' 中目标长时间消失于视野中,LR_CACF 也是唯一准确跟踪目标的算法,同时,目标尺度频繁发生变化,导致预测的矩形框不能很好地包围目标,这也从另一面解释了在 OTB50 和 OTB100 数据集上进行定量评估时,LR_CACF 跟踪器在 AUC 方面获得的增益不太显著,但这并没有造成模型漂移。在序列 Human3 中,目标主要受灰度和低分辨率图像影响,目标特征较少,但是得益于低秩上下文感知约束,LR_CACF 自始至终都稳定跟踪目标。序列 'Biker' 同时面临 IOP、MB 和 OV 等多种挑战,并且多种挑战同时发生,ROT 跟踪器虽然具有重检测机制,但也造成了模型漂移,而 LR_CACF 在面对几种复杂的情况时,都没有丢失目标。

结束语 本文提出了一种基于低秩上下文感知的跟踪算法,在构建目标外观时,目标及其周围的背景信息被同时考虑在内,使得模型对快速运动、平面外旋转等复杂场景比较鲁棒;同时在相邻两帧视频之间施加低秩平滑约束,进一步提高了模型的鉴别性;之后采用 ADMM 框架得到了模型的闭式解,降低了计算复杂度;此外,在滤波模型之外,加入多峰检测和重搜索机制,有效地缓解了遮挡、相似目标干扰等复杂场景下的模型漂移甚至跟踪失败的情况。

最后,在 OTB50、OTB100 和 DTB70 数据集上进一步进行了定性和定量实验,实验结果表明,所提出的 LR_CACF 跟踪器优于其他最先进的跟踪器。基于属性分析的结果表明,LR_CACF 擅长处理快速运动、运动模糊、遮挡和低分辨率等难题。

参考文献

- [1] YANG Y, GU X. Joint Correlation and Attention Based Feature Fusion Network for Accurate Visual Tracking [J]. IEEE Transactions on Image Processing, 2023, 32: 1705-1715.
- [2] JAVED S, DANELLJAN M, KHAN F S, et al. Visual Object Tracking With Discriminative Filters and Siamese Networks: A Survey and Outlook [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(5): 6552-6574.
- [3] ZHU X F, WU X J, XU T, et al. Robust Visual Object Tracking Via Adaptive Attribute-Aware Discriminative Correlation Filters [J]. IEEE Transactions on Multimedia, 2022, 24: 301-312.
- [4] MARVASTI-ZADEH S M, CHENG L, GHANEI-YAKHDAN H, et al. Deep Learning for Visual Tracking: A Comprehensive Survey [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(5): 3943-3968.
- [5] DU S, WANG S. An Overview of Correlation-Filter-Based Object Tracking [J]. IEEE Transactions on Computational Social Systems, 2022, 9(1): 18-31.
- [6] HU W, WANG Q, ZHANG L, et al. SiamMask: A Framework

- for Fast Online Object Tracking and Segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(3): 3072-3089.
- [7] HAN G, SU J, LIU Y, et al. Multi-Stage Visual Tracking With Siamese Anchor-Free Proposal Network [J]. IEEE Transactions on Multimedia, 2023, 25: 430-442.
- [8] 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.
- [9] 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.
- [10] LI Y, ZHU J. A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration [C] // Proceedings of the Computer Vision- ECCV 2014 Workshops. Springer International Publishing, 2015.
- [11] DANELLJAN M, HÄGER G, KHAN F S, et al. Discriminative Scale Space Tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [12] GALOOGAHI H K, FAGG A, LUCEY S. Learning Background-Aware Correlation Filters for Visual Tracking [C] // 2017 IEEE International Conference on Computer Vision (ICCV). 2017: 1144-1152.
- [13] DANELLJAN M, HÄGER G, KHAN F S, et al. Learning Spatially Regularized Correlation Filters for Visual Tracking [C] // Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). 2015: 4310-4318.
- [14] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual Object Tracking using Adaptive Correlation Filters [C] // Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2010: 2544-2550.
- [15] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the Circulant Structure of Tracking-by-Detection with Kernels [C] // Proceedings of the Computer Vision—ECCV. 2012: 702-715.
- [16] LIU T, WANG G, YANG Q. Real-time part - based visual tracking via adaptive correlation filters [C] // Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015: 4902-4912.
- [17] DANELLJAN M, ROBINSON A, KHAN F S, et al. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking [C] // Computer Vision—ECCV. 2016: 472-488.
- [18] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: Efficient Convolution Operators for Tracking [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 6931-6939.
- [19] MUELLER M, SMITH N, GHANEM B. Context-Aware Correlation Filter Tracking [C] // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017: 1387-1395.
- [20] LI F, TIAN C, ZUO W, et al. Learning Spatial-Temporal Regularized Correlation Filters for Visual Tracking [C] // Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018: 4904-4913.
- [21] LI Y, FU C, DING F, et al. AutoTrack: Towards High-Performance Visual Tracking for UAV with Automatic Spatio-Temporal Regularization [C] // Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020: 11920-11929.
- [22] JAIN M, TYAGI A, SUBRAMANYAM A V, et al. Channel Graph Regularized Correlation Filters for Visual Object Tracking [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(2): 715-729.
- [23] ZHANG J, HE Y, WANG S. Learning Adaptive Sparse Spatially-Regularized Correlation Filters for Visual Tracking [J]. IEEE Signal Processing Letters, 2023, 30: 11-15.



SU Yinqiang, born in 1997, Ph.D. His main research interests include visual target tracking and QT-based aviation information processing.



WANG Xuan, born in 1984, Ph.D. associate researcher. His main research interests include airborne photoelectric imaging measurement equipment and so on.

(责任编辑:何杨)