

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2023)02-0458-13

论文引用格式: Tian H D, Zhang J P and Wang Y H. 2023. Sparse constraint and spatial-temporal regularized correlation filter for UAV tracking. Journal of Image and Graphics, 28(02): 0458-0470 (田昊东, 张津浦, 王岳环. 2023. 稀疏约束的时空正则相关滤波无人机视觉跟踪. 中国图象图形学报, 28(02): 0458-0470) [DOI: 10.11834/jig.210611]

稀疏约束的时空正则相关滤波无人机视觉跟踪

田昊东¹, 张津浦¹, 王岳环^{1, 2*}

1. 华中科技大学人工智能与自动化学院, 武汉 430074; 2. 多谱信息处理技术国家级重点实验室, 武汉 430074

摘要: **目的** 基于相关滤波的跟踪算法在无人机(unmanned aerial vehicle, UAV)视觉跟踪领域表现出卓越的性能。现有的相关滤波类跟踪算法从样本区域的所有特征中学习滤波器, 然而某些来自遮挡或形变的特征可能会污染滤波器, 降低模型判别能力。针对此问题, 提出一种稀疏约束的时空正则相关滤波跟踪算法。**方法** 在相关滤波目标函数上施加空间弹性网络约束以自适应地抑制跟踪过程中的干扰特征, 同时集成空间—时间正则相关滤波算法(spatial-temporal regularized correlation filter, STRCF)中的时间正则项以增强滤波器抑制畸变的能力。采用交替方向乘子法(alternating direction method of multipliers, ADMM)将带有约束项的目标函数转化为两个具有闭式解的子问题迭代求局部最优解。此外, 提出一种相关滤波框架通用的加速策略, 根据当前帧的目标位移量, 对检测定位阶段的特征矩阵进行等距离的循环移位, 将其作为在线学习阶段的特征矩阵, 每帧可节省一次训练样本的特征提取操作, 提高跟踪速度。**结果** 在3个UAV数据集上与14种主流跟踪算法进行对比实验, 在DTB70(drone tracking benchmark)数据集中, 平均精确率与平均成功率分别为0.707和0.477, 在所有对比算法中位列第1, 相比较STRCF分别提高了5.8%和4%; 在UAVDT(the unmanned aerial vehicle benchmark: object detection and tracking)数据集中, 平均精确率与平均成功率相比较STRCF分别提高了8.4%和3.8%; 在UAV123_10 fps数据集中, 平均精确率与平均成功率相比较STRCF分别提高了4%和3.3%。同时, 消融实验结果表明, 加速策略在不显著影响跟踪精度($\pm 0.1\%$)的前提下, 可提高跟踪速度约25%, 在单个CPU上的跟踪速度为50帧/s。**结论** 本文算法结合了稀疏约束与时间—空间正则化的优势, 与对比算法相比, 在遮挡、形变等复杂情况下跟踪效果更加鲁棒。

关键词: 无人机跟踪; 相关滤波(CF); 稀疏约束; 时间—空间正则化; 弹性网络; 交替方向乘子法(ADMM); 循环移位

Sparse constraint and spatial-temporal regularized correlation filter for UAV tracking

Tian Haodong¹, Zhang Jinpu¹, Wang Yuehuan^{1, 2*}

1. School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China;

2. National Key Laboratory of Science and Technology on Multi-spectral Information Processing, Wuhan 430074, China

Abstract: **Objective** Correlation filter (CF)-based methods have demonstrated their potential in visual object tracking for unmanned aerial vehicle (UAV) applications. Current discriminative CF trackers can be used to learn a multifaceted feature filter in the sample region. However, more occlusion or deformation-derived features may distort the filter and degrade

收稿日期: 2021-08-12; 修回日期: 2022-01-12; 预印本日期: 2022-01-19

* 通信作者: 王岳环 yuehwang@hust.edu.cn

基金项目: 国家十三五科技预研基金项目(41415020402)

Supported by: Pre-research Foundation of the Thirteenth Five-Year Plan for National Science and Technology of China (41415020402)

the discriminative ability of the model. To mitigate this problem, we develop a novel sparse constraint and spatio-temporal regularized correlation filter to ignore those distractive features adaptively. **Method** By imposing a spatial (bowl-shaped) elastic net constraint on the objective function of the correlation filter, our algorithm can restrict the sparsity of the filter values corresponding to the target region instead of the whole sample region and adaptively suppress the distorted features during tracking. In addition, a temporal regularization term in spatial-temporal regularized correlation filter (STRCF) is integrated to enhance the filter's ability to suppress distortion. Our research treats the object tracking task as a convex optimization problem and provides an efficient global optimization method through alternating direction method of multipliers (ADMM). First, the objective function is required to meet Eckstein-Bertsek condition. Thus, it can converge to the global optimal solution by an unconstrained augmented Lagrange multiplier formulation. Next, ADMM is used to transform the Lagrange multiplier formulation into two sub-problems with closed-form solution. To improve computational efficiency, we convert the sub-problems into the Fourier domain according to Parseval's theorem. Our algorithm can converge quickly within a few iterations. **Result** Several evaluation metrics like center location error and bounding box overlap ratio are used to test and compare the proposed method against other existing methods. The center location error measures the accuracy of the tracking algorithm's estimation for the target location. It computes the average Euclidean distance between the ground truth and the center location of the tracked target in all frames. The center location error can represent the location accuracy of the tracking algorithm. But, the sensitivity of the different size targets to the center location error is different because scale and aspect ratio are not taken into consideration. Another commonly used evaluation metric is the overlap rate, which is defined as the intersection over union between the target box prediction and the ground truth. We compare our approach with several state-of-the-art algorithms on well-known benchmarks, such as DTB70 (drone tracking benchmark), UAVDT (unmanned aerial vehicle benchmark; object detection and tracking) and UAV123_10 fps. The experiment results show that our model outperforms all other methods on DTB70 benchmark. The average accuracy rate and the average success rate are 0.707 and 0.477, which are 5.8% and 4% higher than STRCF. For UAVDT benchmark, the average accuracy rate and the average success rate are 0.72 and 0.494, respectively, which are 8.4% and 3.8% higher than STRCF. For UAV123_10 fps benchmark, the average accuracy rate and average success rate are 0.667 and 0.577, respectively, which are 5% and 3.3% higher than STRCF. Furthermore, an ablation experiment demonstrates that the proposed strategy improves the tracking speed by about 25% without affecting the tracking accuracy, and the running speed can reach 50 frame/s on a single CPU. **Conclusion** Compared with the current popular methods, the proposed sparse constraint and spatio-temporal regularized correlation filter achieves leading performance. Due to the introduction of sparse constraints and spatial-temporal regularization, our algorithm improves the tracking effect and has strong robustness in complex scenes such as occlusion and deformation.

Key words: unmanned aerial vehicle tracking; correlation filter (CF); sparse constraint; spatial-temporal regularization; elastic net; alternating direction method of multipliers (ADMM); cyclic shift

0 引言

随着无人机(unmanned aerial vehicle, UAV)技术的迅猛发展,以无人机为平台的目标识别与跟踪逐渐成为计算机视觉领域的研究热点(王瑶, 2019)。配备了视觉跟踪技术的无人机平台,具有运动灵活、安全性高、体积小和成本低等特点,在导航制导、环境探测和灾难应急等领域得到广泛应用。无人机跟踪任务通常需要持续定位某个特定目标,鲁棒性和实时性是跟踪算法的必然要求。由于无人

机应用场景的复杂性和机载平台计算资源的限制,实现准确、稳定和高效的目标跟踪仍是一项具有挑战性的任务。与一般的地面跟踪任务(如视频监控)相比,无人机平台下的目标跟踪任务面临更严峻的挑战(Fu等, 2022),主要表现在以下方面:1)快速运动。无人机运动灵活、机动性高,机载相机与待跟踪目标之间容易发生剧烈相对运动,对算法快速捕获目标的能力提出了更高要求。同时相机与目标之间的剧烈运动容易导致成像模糊,给跟踪算法的精确定位带来更大挑战。2)视角变化。某些特定应用场景中,无人机可能环绕目标飞行,机载相机

从不同角度捕捉目标的成像姿态,例如车辆的正面和侧面,此时目标外观成像会发生快速变化,若跟踪器无法及时在线学习并更新模型则会导致跟踪失败。3)视觉遮挡。无人机跟踪场景中,目标可能被树木、房屋等外界干扰物部分或完全遮挡,目标特征将被遮挡物污染,容易导致跟踪漂移。4)计算资源稀缺。由于机身体积小,可搭载资源有限,许多无人机仍使用 CPU 作为处理平台,计算能力受限。如何设计跟踪算法,使其在具有较高精度和鲁棒性的前提下,提高计算效率,满足实时性要求,仍面临很大困难。由于以上因素,高效和鲁棒的视觉跟踪算法对于无人机跟踪任务的广泛应用至关重要。

随着卷积神经网络(convolutional neural networks, CNN)的发展,将 CNN 应用于目标跟踪成为研究热点。这类方法通常使用大量离线标记的图像训练用于目标跟踪的 CNN。基于 CNN 的跟踪器具有很高的跟踪精度,但由于网络中卷积运算的高度复杂性,通常在高性能 GPU 上运行,在无人机平台上难以支持;此外, CNN 的离线训练需要海量带标签的预处理图像,这些训练数据的获取需要耗费大量精力;同时,深度网络很容易受到各种噪声的干扰从而降低性能(Yan 等,2020)。因此,基于 CNN 的方法在无人机视觉跟踪领域并非理想的选择。

基于判别式相关滤波(discriminative correlation filter, DCF)的跟踪器具有高速和准确的特点,是无人机视觉跟踪的合适选择。DCF 类跟踪算法最主要的亮点是通过离散傅里叶变换将空间域中循环相关或卷积的复杂计算转换为频域中的元素点乘运算(Bolme 等,2010),这种策略极大提高了 DCF 类跟踪算法的运行速度,大多数 DCF 类跟踪算法在单个 CPU 上的速度达到 30 帧/s 以上,能够满足无人机的实时性要求。

研究人员针对 DCF 类跟踪算法存在的各种问题提出了相应的解决方案,并取得越来越好的结果。针对尺度变化问题, Danelljan 等人(2014a)提出判别式尺度空间(discriminative scale space tracker, DSST)算法,采用位置滤波器和尺度滤波器分别进行目标定位与尺度估计,提高模型的尺度自适应能力。针对边界效应问题, Danelljan 等人(2015)提出空间正则相关滤波算法(spatially regularized discriminative correlation filters, SRDCF),使用 bowl-shaped

结构的空问正则化约束替代传统相关滤波中的常数型 L2 约束,越靠近中心目标位置,惩罚权重越小,反之权重越大,使学习到的滤波器集中在中心区域,同时扩大背景感知区域,提高了跟踪精度。Li 等人(2018)提出空问—时间正则相关滤波算法(spatial-temporal regularized correlation filter, STRCF),在 SRDCF 的目标函数基础上加入时间正则项,抑制相邻两帧之间的滤波器畸变,提高了模型的鲁棒性。针对特征融合问题, Wang 等人(2018)提出多线索相关滤波跟踪算法(multi-cue correlation filter based tracker, MCCT),以 HOG(histogram of oriented gradient)特征(Dalal 和 Triggs, 2005)、CN(color name)特征(Danelljan 等, 2014b)以及 CNN 特征为出发点,采用多专家方式对 3 种特征的所有组合进行打分,在每帧中选择当前最优的特征组合。针对目标外观突变问题, Zheng 等人(2021)提出自适应混合标签的突变感知相关滤波算法(mutation sensitive correlation filter, MSCF),通过一种突变感知因子动态校正训练标签,在外观突变时表现出更好的适应性。

鲁棒视觉跟踪的另一个关键是特征表达的有效性和可靠性。现有的 DCF 类算法通常使用样本区域的所有特征训练滤波器,然而在复杂场景中,某些特征可能会分散注意力,例如目标区域发生遮挡以及形变时,这些干扰因素产生的异常特征会污染模型,降低滤波器的可靠性(Sui 等, 2018)。

本文的主要工作与贡献如下:1)提出一种稀疏约束的相关滤波跟踪算法。通过在目标函数上施加空问弹性网络约束,实现像素级别的特征选择,抑制跟踪过程中的干扰因素。同时,目标函数中集成 STRCF 的时间正则项,提高了模型抑制畸变的能力。最后采用交替方向乘法(alternating direction method of multipliers, ADMM)(Boyd 等, 2011)求解模型。2)提出一种 DCF 框架下通用的加速策略。根据当前帧的目标位移量,对检测定位阶段的特征矩阵进行等距离循环移位,并将其作为在线学习阶段所需要的特征矩阵,使每帧省去一次在线学习阶段的特征提取操作。消融实验表明,该加速策略在不显著影响跟踪精度的前提下,将跟踪速度提高了约 25%。3)在 UAV123_10 fps、DTB70(drone tracking benchmark)和 UAVDT(unmanned aerial vehicle benchmark: object detection and tracking)3 个 UAV

数据集上与 14 种主流算法进行对比实验。结果表明,本文算法相比基线 STRCF,在精确率和成功率上都有显著提高,且优于大多数现有方法,单个 CPU 上的跟踪速度为 50 帧/s。

1 相关工作

1.1 空间—时间正则相关滤波

Li 等人(2018)提出空间—时间正则相关滤波(STRCF),在 SRDCF 的基础上加入时间正则项,抑制相邻两帧的滤波器畸变,防止模型退化。通过求解目标函数,在线学习第 t 帧的最优滤波器 H_t 。该目标函数为

$$\varepsilon(H_t) = \frac{1}{2} \left\| y - \sum_{d=1}^D x_t^d * h_t^d \right\|_2^2 + \frac{1}{2} \sum_{d=1}^D \|w \odot h_t^d\|_2^2 + \frac{\theta}{2} \sum_{d=1}^D \|h_t^d - h_{t-1}^d\|_2^2 \quad (1)$$

式中, $y \in \mathbf{R}^{M \times N}$ 是期望 2 维高斯输出; $x_t^d \in \mathbf{R}^{M \times N}$ ($d = 1, 2, \dots, D$) 是第 t 帧训练样本的特征矩阵, D 表示特征的通道数; $h_t^d, h_{t-1}^d \in \mathbf{R}^{M \times N}$ 表示第 t 和 $t-1$ 帧的滤波器; $*$ 表示卷积操作, $H_t = [h_t^1, h_t^2, \dots, h_t^D]$ 。第 2 项 $\frac{1}{2} \sum_{d=1}^D \|w \odot h_t^d\|_2^2$ 为空间正则项,如图 1 所示,正则化权重 $w \in \mathbf{R}^{M \times N}$ 的分布为 bowl-shaped,对目标中心范围以外的区域惩罚力度更大,期望抑制背景区域的特征对模型的影响,从而可以扩大搜索区域,在目标快速运动时取得更优越的性能; $\odot$ 为逐像素相乘,第 3 项 $\frac{\theta}{2} \sum_{d=1}^D \|h_t^d - h_{t-1}^d\|_2^2$ 为时间正则项,时间正则化权重 θ 为常数,用于惩罚相邻两帧的滤波器畸变。

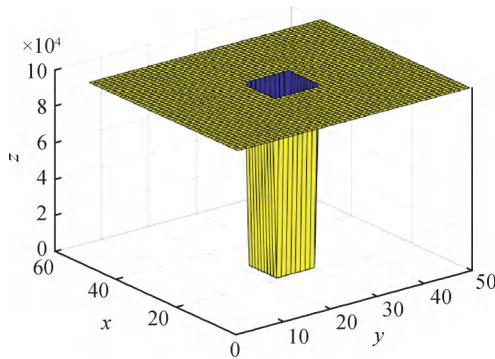


图 1 空间正则化权重

Fig. 1 Spatial regularization weights

空间正则项可以有效缓解传统 DCF 存在的边界效应问题,提高模型的背景感知能力;时间正则项可以适应较大的外观变化,防止滤波器退化。尽管 STRCF 取得了出色的性能,但其仍然存在局限性,具体表现为:使用样本区域的所有特征训练滤波器,使得在复杂的跟踪场景中,由遮挡或形变产生的某些干扰特征会分散模型注意力,使用这些干扰特征训练出的滤波器会大幅降低模型的鲁棒性。

1.2 弹性网络模型

弹性网络是一种使用 L1 和 L2 先验作为正则项的线性回归模型(Zou 和 Hastie, 2005),其目标函数为

$$\varepsilon(w) = \|y - Xw\|_2^2 + \lambda \|w\|_1 + \tau \|w\|_2^2 \quad (2)$$

式中,第 2 项 $\lambda \|w\|_1$ 是 L1 正则项,第 3 项 $\tau \|w\|_2^2$ 是 L2 正则项, λ 与 τ 是常数型的正则化权重。

弹性网络模型包含 Lasso 回归和岭回归约束项,Lasso 回归筛选出相关的特征,并缩减其他无关特征;同时岭回归缩减所有相关性的特征。通过二者的结合,训练的模型像 Lasso 回归一样稀疏,但同时具有岭回归的正则化能力。

弹性网络模型的 L1 和 L2 正则化权重均为常数,限制了整个样本区域中模型的稀疏性。但是在 DCF 跟踪框架中,样本区域就是搜索区域,同时包含了目标以及周围的背景(Peng 和 Lu, 2021)。跟踪过程中的干扰因素通常来自目标区域,例如局部遮挡和目标形变,显然应当更加关注目标区域而非整个样本区域的稀疏性。这意味着式(2)的弹性网络约束并不能直接适用于相关滤波的目标函数。

2 本文算法

2.1 跟踪模型

结合 STRCF 和弹性网络模型的特点,本文构造了稀疏约束的时空正则相关滤波目标函数,定义为

$$\varepsilon(H_t) = \frac{1}{2} \left\| y - \sum_{d=1}^D x_t^d * h_t^d \right\|_2^2 + \sum_{d=1}^D \|w_1 \odot h_t^d\|_1 + \frac{1}{2} \sum_{d=1}^D \|w_2 \odot h_t^d\|_2^2 + \frac{\theta}{2} \sum_{d=1}^D \|h_t^d - h_{t-1}^d\|_2^2 \quad (3)$$

式中,第 2 项 $\sum_{d=1}^D \|w_1 \odot h_t^d\|_1$ 是引入的 L1 空间正则项,用于约束目标区域的稀疏性,其中 $w_1 \in \mathbf{R}^{M \times N}$ 。与弹性网络模型不同,此处的 L1 正则化权重 w_1 是

与空间正则化权重相同的 bowl-shaped 分布,有助于模型更加关注样本中心目标区域的稀疏性,筛选出具有区域特性的判别性特征。此外,目标函数集成为 STRCF 的时间正则项以增强模型抵抗畸变的能力。

图 2 为选自 UAV123 (Mueller 等, 2016) 中的 Car4 遮挡序列,包含车辆从进入到离开遮挡的完整过程。针对遮挡问题,一个有效的方法是在目标区域进行自适应特征选择。在图 2 中,数值小为冷色,数值大为暖色。由于 bowl-shaped 空间正则化的引

入,滤波器的非 0 系数集中在样本中心的目标区域,有助于强调目标区域内的特征信息。图 2(a) 为无遮挡,目标区域的滤波器系数分布正常;图 2(b) 为目标顶部遮挡,顶部区域的滤波器系数明显降低,从而削弱遮挡物对模型的干扰;图 2(c) 为目标底部遮挡,底部区域的滤波器系数明显降低;图 2(d) 为目标驶出遮挡,整个目标区域的滤波器系数恢复到正常水平。图 2 表明,本文算法面对遮挡时具有良好的特征选择能力,可一定程度地抑制遮挡物对模型的污染。

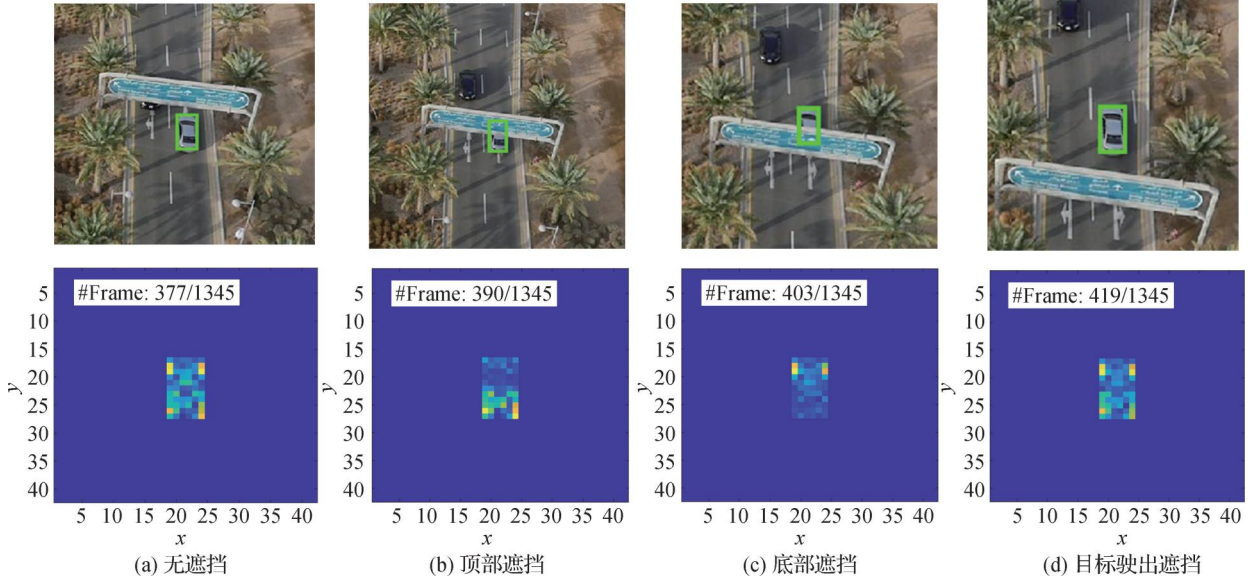


图 2 各种遮挡情况下滤波器的系数分布可视化

Fig. 2 Visualization of coefficients distribution of the filter during occlusion

((a) no occlusion; (b) top occlusion; (c) bottom occlusion; (d) out of occlusion)

图 3 显示了车辆形变时滤波器的系数分布。初始时,跟踪框内仅包含目标,此时滤波器的系数分布较为均匀。随着车辆转弯,由于跟踪框的纵横比固定,跟踪框内不可避免地包含部分背景,但是在弹性网络约束下,跟踪框内背景区域的滤波器系数明显减小,目标区域的滤波器系数保持正常水平,滤波器始终具有较好的目标判别能力。

综上,在目标发生遮挡或形变时,弹性网络约束可以限制干扰区域的滤波器稀疏性,从而削弱该区域的干扰特征对模型的污染,实现像素级别的特征选择,提高模型的鲁棒性。

2.2 优化方法

为了求解目标函数,首先引入辅助变量 $\hat{\mathbf{g}}_t$, $\hat{\mathbf{g}}_t = \hat{\mathbf{h}}_t = \sqrt{T}\mathbf{F}\mathbf{h}_t$, $\hat{\mathbf{G}} = [\hat{\mathbf{g}}_t^1, \hat{\mathbf{g}}_t^2, \dots, \hat{\mathbf{g}}_t^D]$, 其中, T 表示样本区域的总像素个数(宽×高), $\mathbf{F} \in \mathbf{C}^{T \times T}$ 表示离散傅里叶变换矩阵, $\hat{\mathbf{g}}$ 表示向量 $\mathbf{g}$ 的离散傅里叶

变换,此时目标函数式(3)转换到频域,具体为

$$\varepsilon(\mathbf{H}_t, \hat{\mathbf{G}}_t) = \frac{1}{2} \left\| \mathbf{y} - \sum_{d=1}^D \hat{\mathbf{x}}_t^d \odot \hat{\mathbf{g}}_t^d \right\|_2^2 + \sum_{d=1}^D \|\mathbf{w}_1 \odot \mathbf{h}_t^d\|_1 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{w}_2 \odot \mathbf{h}_t^d\|_2^2 + \frac{\theta}{2} \sum_{d=1}^D \|\hat{\mathbf{g}}_t^d - \hat{\mathbf{g}}_{t-1}^d\|_2^2 \quad (4)$$

对于最小化式(4),可以通过 ADMM 迭代求解。式(4)的增广拉格朗日形式为

$$\begin{aligned} L_t(\mathbf{H}_t, \hat{\mathbf{G}}_t, \hat{\mathbf{O}}_t) = & \frac{1}{2} \left\| \mathbf{y} - \sum_{d=1}^D \hat{\mathbf{x}}_t^d \odot \hat{\mathbf{g}}_t^d \right\|_2^2 + \\ & \sum_{d=1}^D \|\mathbf{w}_1 \odot \mathbf{h}_t^d\|_1 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{w}_2 \odot \mathbf{h}_t^d\|_2^2 + \\ & \frac{\theta}{2} \sum_{d=1}^D \|\hat{\mathbf{g}}_t^d - \hat{\mathbf{g}}_{t-1}^d\|_2^2 + \frac{\gamma}{2} \sum_{d=1}^D \|\hat{\mathbf{g}}_t^d - \sqrt{T}\mathbf{F}\mathbf{h}_t^d\|_2^2 + \\ & \sum_{d=1}^D (\hat{\mathbf{g}}_t^d - \sqrt{T}\mathbf{F}\mathbf{h}_t^d) \hat{\mathbf{o}}_t^d \end{aligned} \quad (5)$$

式中, $\hat{\mathbf{O}}_t = [\hat{\mathbf{o}}_t^1, \hat{\mathbf{o}}_t^2, \dots, \hat{\mathbf{o}}_t^D] \in \mathbf{C}^{T \times D}$ 是拉格朗日乘子 $\mathbf{O}_t$ 的离散傅里叶变换形式。 $\gamma \in \mathbf{R}^{1 \times 1}$ 是正则化步

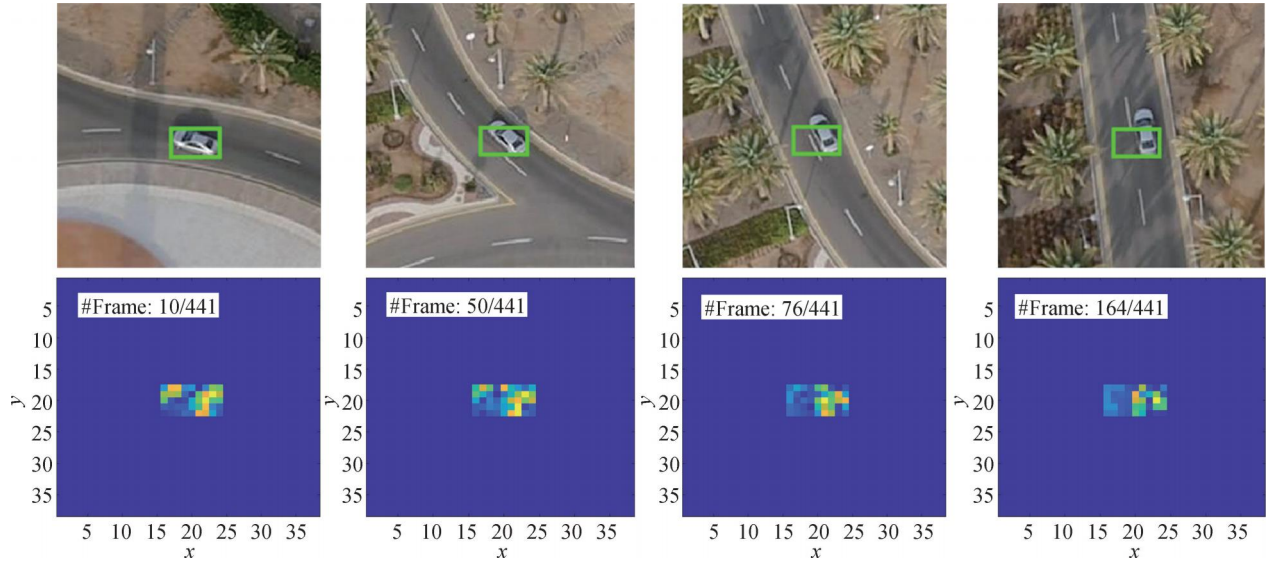


图 3 车辆形变时滤波器的系数分布可视化

Fig. 3 Visualization of filter coefficient distribution during vehicle deformation

长。引入辅助变量 $\mathbf{v}_t$, $\mathbf{v}_t = \frac{\mathbf{o}_t}{\gamma} (\mathbf{V}_t = [\mathbf{v}_t^1, \mathbf{v}_t^2, \dots, \mathbf{v}_t^D])$, 式(5)可转化为

$$L_t(\mathbf{H}_t, \hat{\mathbf{G}}_t, \hat{\mathbf{V}}_t) = \frac{1}{2} \left\| \mathbf{y} - \sum_{d=1}^D \hat{\mathbf{x}}_t^d \odot \hat{\mathbf{g}}_t^d \right\|_2^2 + \sum_{d=1}^D \|\mathbf{w}_1 \odot \mathbf{h}_t^d\|_1 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{w}_2 \odot \mathbf{h}_t^d\|_2^2 + \frac{\theta}{2} \sum_{d=1}^D \|\hat{\mathbf{g}}_t^d - \hat{\mathbf{g}}_{t-1}^d\|_2^2 + \frac{\gamma}{2} \sum_{d=1}^D \|\hat{\mathbf{g}}_t^d - \sqrt{T} \mathbf{F} \mathbf{h}_t^d + \hat{\mathbf{v}}_t^d\|_2^2 \quad (6)$$

式(6)可以拆分为具有闭式解的子问题 $\hat{\mathbf{G}}$ 和子问题 $\mathbf{H}$, 分别在频域和空域求解。

1) 子问题 $\hat{\mathbf{G}}$ 。给定 $\mathbf{H}_t, \hat{\mathbf{V}}_t$, 最优的 $\hat{\mathbf{G}}^*$ 为

$$\hat{\mathbf{G}}^* = \arg \min_{\hat{\mathbf{G}}} \left\{ \frac{1}{2} \left\| \hat{\mathbf{y}} - \sum_{d=1}^D \hat{\mathbf{x}}_t^d \odot \hat{\mathbf{g}}_t^d \right\|_2^2 + \frac{\theta}{2} \sum_{d=1}^D \|\hat{\mathbf{g}}_t^d - \hat{\mathbf{g}}_{t-1}^d\|_2^2 + \frac{\gamma}{2} \sum_{d=1}^D \|\hat{\mathbf{g}}_t^d - \sqrt{T} \mathbf{F} \mathbf{h}_t^d + \hat{\mathbf{v}}_t^d\|_2^2 \right\} \quad (7)$$

由于 D 维通道在计算时相互独立, 因此在每个像素的所有通道中对 $\hat{\mathbf{x}}_t$ 进行采样, 式(7)可转换为

$$\Gamma_j^*(\hat{\mathbf{G}}) = \arg \min_{\Gamma_j(\hat{\mathbf{G}})} \left\{ \frac{1}{2} \left\| \hat{\mathbf{y}} - \Gamma_j(\hat{\mathbf{X}}_t)^\top \Gamma_j(\hat{\mathbf{G}}_t) \right\|_2^2 + \frac{\gamma}{2} \left\| \Gamma_j(\hat{\mathbf{G}}_t) - \Gamma_j(\sqrt{T} \mathbf{F} \mathbf{H}_t) + \Gamma_j(\hat{\mathbf{V}}_t) \right\|_2^2 + \frac{\theta}{2} \left\| \Gamma_j(\hat{\mathbf{G}}_t) - \Gamma_j(\hat{\mathbf{G}}_{t-1}) \right\|_2^2 \right\} \quad (8)$$

式中, $\Gamma_j(\hat{\mathbf{X}}_t) \in \mathbb{C}^{D \times 1}$ 表示样本 $\hat{\mathbf{X}}$ 上第 j 个像素的

D 维特征向量, $j=1, 2, \dots, T$ 。对式(8)求导, 并令导数为 0, 得到 $\Gamma_j^*(\hat{\mathbf{G}})$ 的闭式解为

$$\Gamma_j^*(\hat{\mathbf{G}}) = (\Gamma_j(\hat{\mathbf{X}}_t) \Gamma_j(\hat{\mathbf{X}}_t)^\top + (\gamma + \theta) \mathbf{I})^{-1} \boldsymbol{\rho} \quad (9)$$

式中, 向量 $\boldsymbol{\rho}$ 的形式为

$$\boldsymbol{\rho} = \Gamma_j(\hat{\mathbf{X}}_t) \hat{\mathbf{y}}_j + \theta \Gamma_j(\hat{\mathbf{G}}_{t-1}) - \gamma \Gamma_j(\hat{\mathbf{V}}_t) + \gamma \Gamma_j(\sqrt{T} \mathbf{F} \mathbf{H}_t) \quad (10)$$

式(9)中仍含有时间复杂度很高的求逆运算, 使用 Sherman-Morrison 公式 (Sherman 和 Morrison, 1950) 对其进行转换, 最终得到 $\Gamma_j^*(\hat{\mathbf{G}})$ 的闭式解为

$$\Gamma_j^*(\hat{\mathbf{G}}) = \frac{1}{\gamma + \theta} \left(\mathbf{I} - \frac{\Gamma_j(\hat{\mathbf{X}}_t) \Gamma_j(\hat{\mathbf{X}}_t)^\top}{\gamma + \theta + \Gamma_j(\hat{\mathbf{X}}_t) \Gamma_j(\hat{\mathbf{X}}_t)^\top} \right) \boldsymbol{\rho} \quad (11)$$

2) 子问题 $\mathbf{H}$ 。给定 $\hat{\mathbf{G}}_t, \hat{\mathbf{V}}_t$, 最优的 $\mathbf{h}^{d*}$ 为

$$\mathbf{h}^{d*} = \arg \min_{\mathbf{h}} \left\| \mathbf{w}_1 \odot \mathbf{h}_t^d \right\|_1 + \frac{1}{2} \left\| \mathbf{w}_2 \odot \mathbf{h}_t^d \right\|_2^2 + \frac{\gamma}{2} \sum_{d=1}^D \left\| \hat{\mathbf{g}}_t^d - \sqrt{T} \mathbf{F} \mathbf{h}_t^d + \hat{\mathbf{v}}_t^d \right\|_2^2 \quad (12)$$

由于给定了 $\hat{\mathbf{g}}_t^d$ 和 $\hat{\mathbf{v}}_t^d$, 式(12)属于 L1 范数正则化的最小二乘问题, 可以通过阈值收缩法 (Beck 和 Teboulle, 2009) 求解, 式(12)具有全局唯一解。具体为

$$\mathbf{h}^{d*} = (\mathbf{W}_2 \mathbf{W}_2 + \gamma \mathbf{I})^{-1} \eta(\gamma T(\mathbf{g}^d + \mathbf{v}^d), \mathbf{w}_1) = \frac{\eta(\gamma T(\mathbf{g}^d + \mathbf{v}^d), \mathbf{w}_1)}{\mathbf{w}_2 \odot \mathbf{w}_2 + \gamma T} \quad (13)$$

式中, $\mathbf{W}_2 = \text{diag}(\mathbf{w}_2) \in \mathbb{R}^{T \times T}$ 表示对角矩阵。 η 是软

阈值操作,定义为

$$\eta(x, \lambda) = \text{sign}(x) \max(|x| - \lambda, 0) \quad (14)$$

3) 拉格朗日乘子更新。求解出以上两个子问题后,更新拉格朗日乘子,具体为

$$\hat{\mathbf{V}}^{i+1} = \hat{\mathbf{V}}^i + \gamma^i (\hat{\mathbf{G}}^{i+1} - \hat{\mathbf{H}}^{i+1}) \quad (15)$$

式中, i 和 $i+1$ 表示迭代次数。

正则化步长 γ (初始值为 1) 更新式为

$$\gamma^{i+1} = \min(\gamma_{\max}, \beta \gamma^i) \quad (16)$$

式中, $\beta = 10$, $\gamma_{\max} = 10\,000$ 。

通过迭代求解子问题 $\hat{\mathbf{G}}$ 和子问题 $\hat{\mathbf{H}}$, 并且更新拉格朗日乘子, 可以有效优化目标函数, 获得当前帧的最优滤波器 $\hat{\mathbf{G}}$, 用于下一帧的目标定位。

2.3 目标定位方法

使用最优滤波器 $\hat{\mathbf{G}}$ 计算响应图 $\mathbf{R}_t$, 具体为

$$\mathbf{R}_t = F^{-1} \sum_{d=1}^D (\hat{\mathbf{z}}_t^d \odot \hat{\mathbf{g}}_{t-1}^d) \quad (17)$$

式中, $\mathbf{R}_t$ 表示第 t 帧的响应图, F^{-1} 表示傅里叶逆变换 (inverse Fourier transform, IFT) 操作, $\hat{\mathbf{z}}_t^d$ 表示检测定位阶段样本特征矩阵的频域表示。响应图 $\mathbf{R}_t$ 中最大值所在的位置即为第 t 帧的目标中心位置。

针对尺度估计问题, 借用 fDSST (fast DSST) (Danelljan 等, 2017a) 的方法, 在原有位置滤波器的基础上, 增加一个尺度滤波器。通过位置滤波器确定目标中心位置后, 在相同的中心位置上, 提取多个尺度等级的样本用于训练尺度滤波器, 根据尺度滤波器计算尺度相关得分的最大值, 从而确定当前帧的最优尺度。

2.4 加速策略

本文提出一种 DCF 框架通用的加速策略, 通过对检测定位阶段的样本特征矩阵进行循环移位, 并

将其作为训练阶段的样本特征矩阵, 可节省一次特征提取操作, 大幅度提高跟踪速度。

在 DCF 类跟踪框架中, 每一帧的处理流程可分为检测定位与在线学习两个阶段。其中, 检测定位阶段旨在以上一帧的目标位置 pos_{t-1} 为中心, 提取检测定位样本 (图 4 虚线绿框) 的特征矩阵, 与滤波器相关计算得到当前帧的目标位置 pos_t ; 在线学习阶段旨在以计算出的新目标位置 pos_t 为中心, 提取在线学习所需样本 (图 4 虚线红框) 的特征矩阵, 进而求解出新一帧的滤波器模型。



图 4 目标检测样本 (绿框) 和在线学习样本 (红框)

Fig. 4 Target detection sample (green box) and online learning sample (red box)

两个阶段共需要进行两次特征提取与离散傅里叶变换操作。两个样本的中心偏移量 (dx, dy) 为

$$\begin{cases} dx = pos_t(x) - pos_{t-1}(x) \\ dy = pos_t(y) - pos_{t-1}(y) \end{cases} \quad (18)$$

样本图像块如图 5 所示。其中, 图 5(a) 为检测定位阶段的样本, 图 5(b) 为图 5(a) 经过 (dx, dy) 的循环移位得到的移位样本, 图 5(c) 为真实的在线学习所需样本。可以看出, 图 5(b) 与图 5(c) 仅在边界存在差异, 图 5(b) 左边界失真。在 DCF 框架



(a) 检测样本

(b) 移位样本

(c) 真实样本

图 5 样本图像块

Fig. 5 Samples of image block ((a) detection sample; (b) shifted sample; (c) real sample)

中,边界区域受到余弦窗与空间正则项的抑制,对滤波器的判别能力几乎不构成影响,因此可以使用移位样本(图 5(b))代替真实样本(图 5(c))进行滤波器的在线学习。同时,由于密集采样提取的特征值与像素点具有位置对应关系,因此直接对检测样本(图 5(a))的特征矩阵进行 (dx, dy) 的循环移位,可得移位样本(图 5(b))的特征矩阵,并将其用于后续的在线学习。

由于滤波器学习阶段使用样本特征的频域表示,根据傅里叶变换的时移特性,时域信号的循环移位对应频域信号的相位旋转,因此可将检测样本(图 5(a))的频域特征值乘以对应的虚指数,直接在频域进行特征转换,进一步提高效率。消融实验结果表明,在对跟踪精度没有显著影响($\pm 0.1\%$)的前提下,该加速策略可将跟踪速度提高 25%。

3 实验结果与分析

3.1 参数设置与实验环境

图像特征选择手工特征 HOG 和 CN,模型的超参数参考 STRCF(Li 等,2018),目标函数中的 L1 和 L2 正则化权重均为 bowl-shaped 分布,其中 L2 正则项参数 w_2 中心区域设置为 1×10^{-3} ,四周区域设置为 1×10^5 ;时间正则化权重 θ 设为 13;正则化步长 γ 初始为 1。针对本文的空间弹性网络约束项 $\sum_{d=1}^D \|\mathbf{w}_1 \odot \mathbf{h}_t^d\|_1 + \frac{1}{2} \sum_{d=1}^D \|\mathbf{w}_2 \odot \mathbf{h}_t^d\|_2^2$,设置 L1 正则化权重 $\mathbf{w}_1 = \frac{1}{2} \mathbf{w}_2 \odot \mathbf{w}_2$,使 L1 与 L2 的约束力度对等。若 L1 正则化权重设置为 0,则退化为 STRCF。

实验软件环境为 MATLAB R2017b,硬件环境为 Inter(R) Core(TM) i7-7700HQ CPU,2.80 GHz 主频,8 GB 内存。

3.2 UAV 数据集介绍

为了进行全面评估,实验在 3 个具有挑战性和权威性的 UAV 数据集上进行,分别是 UAV123_10 fps(Mueller 等,2016)、DTB70(Li 和 Yeung,2017)和 UAVDT(Du 等,2018)数据集,共 243 个视频序列。

UAV123 数据集全部由无人机在空中拍摄,视角变化大,包含 123 个视频序列,超过 11 万帧图像。UAV123_10 fps 是从原序列中以抽帧的方式创建的 10 帧/s 的数据集,因此跟踪过程中目标的位置移

动、姿态变化更大,带来了更多挑战。DTB70 数据集由 70 个困难的无人机视频序列组成,主要针对无人机拍摄过程中剧烈运动以及目标的大小和纵横比变化问题。UAVDT 数据集主要针对各种情况下的车辆跟踪,例如飞行高度、摄像机视角和车辆遮挡等。

3.3 实验结果分析

为了验证本文算法的先进性,在 3 个 UAV 数据集上与 14 种主流算法进行对比实验,分别为 MRCF(multi-regularized correlation filter)(Ye 等,2022)、MSCF(mutation sensitive correlation filter)(Zheng 等,2021)、ARCF(aberrance repressed correlation filter)(Huang 等,2019)、STRCF(Li 等,2018)、MCCT(multi-cue correlation filter based tracker)(Wang 等,2018)、ECO-HC(efficient convolution operators—hand-crafted feature version)(Danelljan 等,2017b)、BACF(background-aware correlation filters)(Galoogahi 等,2017)、fDSST(Danelljan 等,2017a)、Staple(fast discriminative scale space tracker)(Bertinetto 等,2016)、SDRCF(Danelljan 等,2015)、LCT(long-term correlation tracking)(Ma 等,2015a)、DSST(Danelljan 等,2015)、KCF(kernelized correlation filter)(Henriques 等,2015)、SAMF(scale adaptive with multiple features tracker)(Li 和 Zhu,2014)。考虑到无人机平台的实现能力,所有跟踪算法均采用手工特征,确保在单个 CPU 上的效率,一些采用 DCF 结构但具有 CNN 特征的跟踪器,例如 ASRCF(adaptive spatially-regularized correlation filters)(Dai 等,2019)不属于本文研究范畴。对比算法均使用原文献提供的开源代码或结果。

3.3.1 与 STRCF 算法定性比较

本文方法与 STRCF 算法对包含视角变化、快速运动、低分辨率、快速形变以及遮挡等挑战的 4 个视频序列的跟踪结果如图 6 所示。可以看出,本文方法的跟踪性能明显优于 STRCF。

1) 视角变化。由于无人机拍摄时自身发生旋转,导致目标成像姿态快速变化。快速旋转时目标的梯度信息变化剧烈,此时 HOG 特征的表征能力欠佳。本文算法采用 HOG + CN 组合特征,在快速旋转这种梯度敏感的场景下,弹性网络约束可以减弱 HOG 特征的影响,加强 CN 颜色特征对模型的贡献,使模型具有更好的鲁棒性。如图 6(a)所示,本文算法相比 STRCF 具有更高的定位精度。



图6 本文方法与 STRCF 算法对比

Fig. 6 Comparison between STRCF and ours ((a) Car8; (b) ChasingDrones; (c) Motor2; (d) Horse)

2) 快速运动与低分辨率。从图 6(b) 可以看出, 跟踪目标的尺寸小、分辨率低, 在目标特征信息较少的情况下, 本文算法依然可以准确跟踪。

3) 快速形变。在图 6(c) 杂技表演序列中, 表演者骑车在空中翻转, 自身发生快速形变, STRCF 无法适应这种快速形变而跟踪失败, 本文算法通过弹性网络约束提高特征的可靠性, 增强了模型的判别能力, 能够更好地定位目标。

4) 视觉遮挡。遮挡是视觉目标跟踪的常见问题。在图 6(d) Horse 视频序列中, 在目标进入遮挡、完全遮挡和离开遮挡的过程中, STRCF 模型被树木遮挡物污染, 跟踪漂移。本文算法通过稀疏约束, 自适应地忽略遮挡区域的干扰特征, 防止模型学习到错误信息, 能够在一定程度上抵抗短期遮挡。

上述分析表明, 本文算法与 STRCF 相比在视角变化、快速运动、快速形变和视觉遮挡等复杂场景下

具有更好的鲁棒性与有效性。

3.3.2 与 14 种算法定量比较

本文方法与 14 种跟踪算法在 3 个 UAV 数据集上的精确率和成功率对比结果如图 7、图 8 和表 1 所示。跟踪性能最好的 6 种算法的精确率、成功率 (Wu 等, 2015) 与跟踪速率 (frames per second, FPS) 如表 2 所示。

在 UAV123_10 fps 数据集上, 本文算法的精确率为 0.667, 在所有算法中位列第 1, 相比 STRCF 算法提高了 4%。本文算法的成功率为 0.577, 在所有算法中位列第 2, 相比 STRCF 算法提高了 3.3%。在 DTB70 数据集上, 本文算法的精确率与成功率分别为 0.704 和 0.477, 在所有算法中均位列第 1, 相比 STRCF 算法分别提高了 5.8% 和 4%。在 UAVDT 数据集上, 本文算法的精确率为 0.720, 在所有算法中位列第 4, 相比 STRCF 算法提高了 8.4%。本文

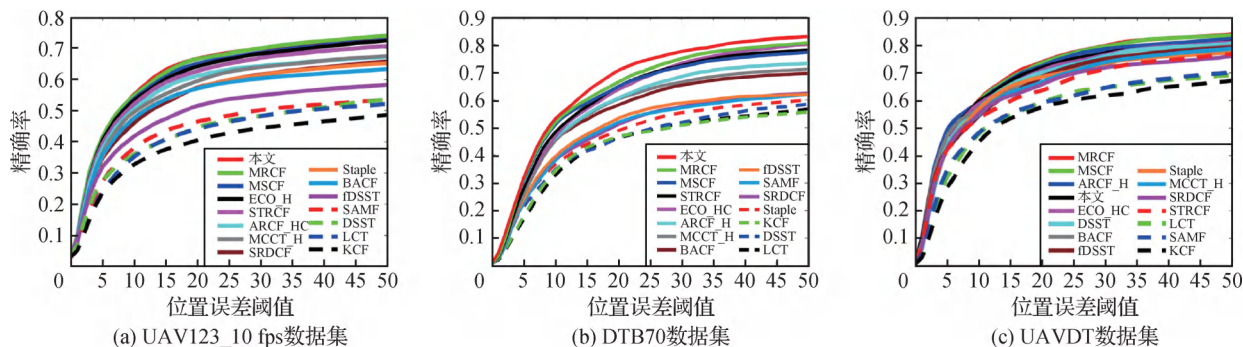


图 7 15 种跟踪器在 UAV 基准上的精确率曲线

Fig. 7 Curves of precision for 15 trackers on UAV benchmark

((a) UAV123_10 fps dataset; (b) DTB70 dataset; (c) UAVDT dataset)

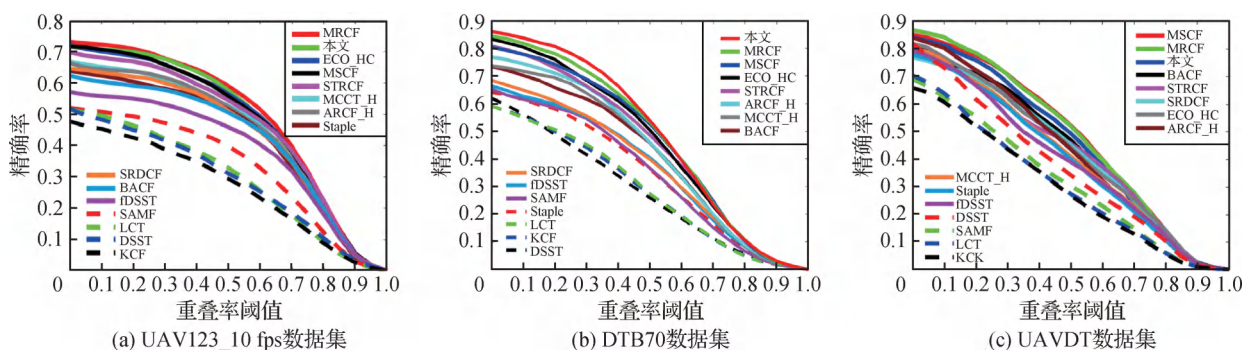


图 8 15 种跟踪器在 UAV 基准上的成功率曲线

Fig. 8 Curves of success rate for 15 trackers on UAV benchmark

((a) UAV123_10 fps dataset; (b) DTB70 dataset; (c) UAVDT dataset)

表 1 15 种跟踪器在 UAV 基准上的精确率和成功率结果

Table 1 Results of precision and success rates for 15 trackers on UAV benchmark

跟踪器	精确率			成功率		
	UAV123_10 fps	DTB70	UAVDT	UAV123_10 fps	DTB70	UAVDT
本文	0.667	0.707	0.720	0.577	0.477	0.494
MRCF	0.666	0.666	0.740	0.591	0.466	0.522
MSCF	0.651	0.649	0.737	0.558	0.450	0.524
ECO_HC	0.643	0.638	0.709	0.565	0.449	0.443
STRCF	0.627	0.649	0.636	0.544	0.437	0.456
ARCF_H	0.612	0.607	0.725	0.524	0.416	0.437
MCCT_H	0.596	0.604	0.668	0.526	0.405	0.411
SRDCF	0.575	0.512	0.659	0.511	0.363	0.452
Staple	0.573	0.489	0.685	0.515	0.336	0.402
BACF	0.572	0.581	0.686	0.506	0.398	0.469
fDSST	0.516	0.534	0.686	0.459	0.357	0.399
SAMF	0.465	0.520	0.591	0.398	0.339	0.303
DSST	0.448	0.463	0.702	0.311	0.276	0.343
LCT	0.442	0.462	0.594	0.329	0.283	0.275
KCF	0.405	0.468	0.575	0.294	0.276	0.269

表 2 跟踪性能最好的 6 种跟踪器在 UAV123_10 fps、DTB70 和 UAVDT 数据集上的精确率、成功率和 FPS 结果
Table 2 Results of precision, success rates and FPS for top six trackers on UAV123_10 fps, DTB70 and UAVDT datasets

方法	UAV123_10 fps			DTB70			UAVDT		
	精确率	成功率	FPS/(帧/s)	精确率	成功率	FPS/(帧/s)	精确率	成功率	FPS/(帧/s)
STRCF	0.627	0.544	28	0.649	0.437	22	0.636	0.456	35
MRCF	<u>0.666</u>	0.591	37	<u>0.666</u>	<u>0.466</u>	35	0.740	<u>0.522</u>	53
MSCF	0.651	0.558	31	0.649	0.450	27	<u>0.737</u>	0.524	37
ECO-HC	0.643	0.565	55	0.638	0.449	<u>46</u>	0.709	0.443	60
ARCF-H	0.612	0.524	46	0.607	0.416	47	0.725	0.437	51
本文	0.667	<u>0.577</u>	<u>50</u>	0.707	0.477	<u>46</u>	0.72	0.494	<u>59</u>

注:加粗字体表示各列最优结果,下划线字体表示各列次优结果。

算法的成功率为 0.494,在所有算法中位列第 3,相比 STRCF 算法提高了 3.8%。通过引入弹性网络约束,本文算法能够自适应地抑制异常情况下的干扰特征,提高模型在复杂场景下的适应能力,使得精度率与成功率均明显优于 STRCF。

与 14 种当前主流算法相比,本文方法在精度与速度上均优于大多数现有方法,证明了本文方法在无人机视觉跟踪领域的优越性。

3.4 消融实验

根据 2.4 节的分析可知,对检测定位阶段的样本进行循环移位会导致移位样本存在边界失真现象,且相邻两帧之间目标的位置移动越剧烈,边界失真现象越显著。为了更好地验证本文提出的稀疏约束与加速策略对跟踪精度与跟踪速度带来的具体影响,采用目标位置移动、姿态变化更剧烈的 UAV123_10 fps 数据集进行消融实验。

消融实验中的 4 种方法均使用 HOG + CN 手工特征。Baseline 在原始 DCF 框架中加入空间正则项与时间正则项,未使用弹性网络 (acceleration strategy, AS) 和加速策略 (elastic net, EN); Baseline + AS 表示在 Baseline 基础上使用加速策略; Baseline + EN 表示在 Baseline 基础上引入空间弹性网络对模型进行区域性稀疏约束;本文算法包含区域稀疏约束项与加速策略。消融实验结果如表 3 所示。可以看出,使用弹性网络约束的 Baseline + EN 和本文跟踪器,对比未使用弹性网络约束的 Baseline 与 Baseline + AS 跟踪器,精确率和成功率均有提高。使用加速策略的 Baseline + AS 和本文跟踪器,对比未使用加速策略的 Baseline 和 Baseline + EN 跟踪器,跟踪速率(帧/s)明显提高。此外,本文

跟踪器与 Baseline + EN 跟踪器相比,额外引入加速策略,在精确率和成功率仅下降 0.1% 的情况下,跟踪速率(帧/s)提高了约 25%。

表 3 消融实验
Table 3 Ablation experiments

跟踪器	精确率	成功率	跟踪速率/(帧/s)
Baseline	0.646	0.558	39
Baseline + AS	0.648	0.548	51
Baseline + EN	0.668	0.578	40
Baseline + EN + AS (本文)	0.667	0.577	50

注:加粗字体表示各列最优结果。

需要注意的是, UAV123_10 fps 数据集采用 3 帧抽 1 的方式创建,目标相邻两帧之间的位置移动更大,加速策略中的循环移位操作会带来更多的边界失真,但得益于相关滤波框架中余弦窗与空间正则项的作用,边界对模型的影响受到极大的抑制。消融实验结果表明,加速策略大幅提高了跟踪速度,同时没有显著降低跟踪精确率和成功率。综上,本文提出的弹性网络约束与加速策略可以显著改善 DCF 类跟踪算法的性能。

4 结 论

针对无人机视觉跟踪任务对准确性、鲁棒性和实时性的高要求,本文提出一种稀疏约束的时空正则相关滤波跟踪算法,通过将弹性网络约束与空间正则项结合,使相关滤波器在遮挡、形变和视角变化等复杂场景下能够自适应地筛选具有区域特性的判

别性特征;同时目标函数中集成了时间正则项以缓解畸变导致的模型退化;采用 ADMM 方法将带等式约束的目标函数转化为两个具有闭式解的子问题,迭代求取局部最优解。此外,提出一种 DCF 框架下通用的加速策略,通过在频域计算检测定位阶段的样本特征矩阵的循环移位,高效地获取滤波器在线学习阶段的样本特征矩阵,在不显著影响跟踪精度 ($\pm 0.1\%$) 的前提下,跟踪速度提高约 25%,本文算法在单个 CPU 上的速度约 50 帧/s。

在 UAV123_10 fps、DTB70 和 UAVDT 这 3 个具有挑战性与权威性的 UAV 数据集上与 14 种主流算法进行对比实验。本文算法与基线 STRCF 相比,在精确率、成功率以及跟踪速度上均具有明显优势;与其他主流的视觉跟踪算法相比,本文算法在精度与速度上同样优于大多数 DCF 类跟踪算法,证明了本文算法在无人机视觉跟踪领域的优越性。

现有跟踪方法普遍忽视了多通道特征的差异性,对所有特征通道一视同仁,从而限制了多通道特征在不同场景下的适应能力。下一步工作将考虑加入场景感知模块,通过判断不同场景下各个特征通道的可靠性,自适应地调整各个特征通道的权重分布,提高模型在复杂场景下的适应能力。

参考文献 (References)

- Beck A and Teboulle M. 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1): 183-202 [DOI: 10.1137/080716542]
- Bertinetto L, Valmadre J, Golodetz S, Miksik O and Torr P H S. 2016. Staple: complementary learners for real-time tracking//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA; IEEE: 1401-1409 [DOI: 10.1109/CVPR.2016.156]
- Bolme D S, Beveridge J R, Draper B A and Lui Y M. 2010. Visual object tracking using adaptive correlation filters//*Proceedings of 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Francisco, USA; IEEE: 2544-2550 [DOI: 10.1109/CVPR.2010.5539960]
- Boyd S, Parikh N, Chu E, Peleato B and Eckstein J. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1): 1-122 [DOI: 10.1561/22000000016]
- Dai K N, Wang D, Lu H C, Sun C and Li J H. 2019. Visual tracking via adaptive spatially-regularized correlation filters//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Long Beach, USA; IEEE: 4665-4674 [DOI: 10.1109/CVPR.2019.00480]
- Dalal N and Triggs B. 2005. Histograms of oriented gradients for human detection//*Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. San Diego, USA; IEEE: 886-893 [DOI: 10.1109/CVPR.2005.177]
- Danelljan M, Bhat G, Khan F S and Felsberg M. 2017b. ECO: efficient convolution operators for tracking//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, USA; IEEE: 6931-6939 [DOI: 10.1109/CVPR.2017.733]
- Danelljan M, Häger G, Khan F S and Felsberg M. 2014a. Accurate scale estimation for robust visual tracking//*Proceedings of 2014 British Machine Vision Conference*. Nottingham, UK; BMVA Press: 1-65 [DOI: 10.5244/C.28.65]
- Danelljan M, Häger G, Khan F S and Felsberg M. 2015. Learning spatially regularized correlation filters for visual tracking//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile; IEEE: 4310-4318 [DOI: 10.1109/ICCV.2015.490]
- Danelljan M, Häger G, Khan F S and Felsberg M. 2017a. Discriminative scale space tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8): 1561-1575 [DOI: 10.1109/TPAMI.2016.2609928]
- Danelljan M, Khan F S, Felsberg M and van de Weijer J. 2014b. Adaptive color attributes for real-time visual tracking//*Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA; IEEE: 1090-1097 [DOI: 10.1109/CVPR.2014.143]
- Du D W, Qi Y K, Yu H Y, Yang Y F, Duan K W, Li G R, Zhang W G, Huang Q M and Tian Q. 2018. The unmanned aerial vehicle benchmark: object detection and tracking//*Proceedings of the 15th European Conference Computer Vision*. Munich, Germany; Springer: 375-391 [DOI: 10.1007/978-3-030-01249-6_23]
- Fu C H, Li B W, Ding F Q, Lin F L and Lu G. 2022. Correlation filters for unmanned aerial vehicle-based aerial tracking: a review and experimental evaluation. *IEEE Geoscience and Remote Sensing Magazine*, 10(1): 125-160 [DOI: 10.1109/MGRS.2021.3072992]
- Galoogahi H K, Fagg A and Lucey S. 2017. Learning background-aware correlation filters for visual tracking//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy; IEEE: 1144-1152 [DOI: 10.1109/ICCV.2017.129]
- Henriques J F, Caseiro R, Martins P and Batista J. 2015. High-speed tracking with Kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3): 583-596 [DOI: 10.1109/TPAMI.2014.2345390]
- Huang Z Y, Fu C H, Li Y M, Lin F L and Lu P. 2019. Learning aberrance repressed correlation filters for real-time UAV tracking//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): 2891-2900 [DOI: 10.1109/ICCV.

- 2019.00298]
- Li F, Tian C, Zuo W M, Zhang L and Yang M H. 2018. Learning spatial-temporal regularized correlation filters for visual tracking//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 4904-4913 [DOI: 10.1109/CVPR.2018.00515]
- Li S Y and Yeung D Y. 2017. Visual object tracking for unmanned aerial vehicles: a benchmark and new motion models//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA; AAAI Press: 4140-4146
- Li Y and Zhu J K. 2014. A scale adaptive kernel correlation filter tracker with feature integration//Proceedings of 2014 Computer Vision. Zurich, Switzerland; Springer: 254-265 [DOI: 10.1007/978-3-319-16181-5_18]
- Ma C, Yang X K, Zhang C Y and Yang M H. 2015a. Long-term correlation tracking//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA; IEEE: 5388-5396 [DOI: 10.1109/CVPR.2015.7299177]
- Mueller M, Smith N and Ghanem B. 2016. A benchmark and simulator for UAV tracking//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands; Springer: 445-461 [DOI: 10.1007/978-3-319-46448-0_27]
- Peng Z and Lu X J. 2021. Learning region sparse constraint correlation filter for tracking. Signal Processing: Image Communication, 90; #116042 [DOI: 10.1016/j.image.2020.116042]
- Sherman J and Morrison W J. 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. The Annals of Mathematical Statistics, 21(1): 124-127 [DOI: 10.1214/aoms/1177729893]
- Sui Y, Wang G H and Zhang L. 2018. Correlation filter learning toward peak strength for visual tracking. IEEE Transactions on Cybernetics, 48(4): 1290-1303 [DOI: 10.1109/TCYB.2017.2690860]
- Wang N, Zhou W G, Tian Q, Hong R C, Wang M and Li H Q. 2018. Multi-cue correlation filters for robust visual tracking//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA; IEEE: 4844-4853 [DOI: 10.1109/CVPR.2018.00509]
- Wang Y. 2019. Vision-Based UAV Target Recognition and Tracking. Harbin: Harbin Institute of Technology (王瑶. 2019. 基于视觉的无人机目标识别及跟踪. 哈尔滨: 哈尔滨工业大学)
- Wu Y, Lim J and Yang M H. 2015. Object tracking benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9): 1834-1848 [DOI: 10.1109/TPAMI.2014.2388226]
- Yan B, Wang D, Lu H C and Yang X Y. 2020. Cooling-shrinking attack: blinding the tracker with imperceptible noises//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA; IEEE: 987-996 [DOI: 10.1109/CVPR42600.2020.00107]
- Ye J J, Fu C H, Lin F L, Ding F Q, An S and Lu G. 2022. Multi-regularized correlation filter for UAV tracking and self-localization. IEEE Transactions on Industrial Electronics, 69(6): 6004-6014 [DOI: 10.1109/TIE.2021.3088366]
- Zheng G Z, Fu C H, Ye J J, Lin F L and Ding F Q. 2021. Mutation sensitive correlation filter for real-time UAV tracking with adaptive hybrid label//Proceedings of 2021 IEEE International Conference on Robotics and Automation. Xi'an, China; IEEE: 503-509 [DOI: 10.1109/ICRA48506.2021.9561931]
- Zou H and Hastie T. 2005. Addendum: regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B (Statistical Methodology), 67(5): #768 [DOI: 10.1111/j.1467-9868.2005.00527.x]

作者简介

田昊东,男,硕士研究生,主要研究方向为计算机视觉与模式识别、目标跟踪。E-mail: 18821704039@163.com

王岳环,通信作者,男,教授,主要研究方向为计算机视觉、实时自动目标识别、嵌入式系统。

E-mail: yuehwang@hust.edu.cn

张津浦,男,博士研究生,主要研究方向为计算机视觉与模式识别。E-mail: zjphust@gmail.com