

采用多边形质心的相关滤波跟踪位置校正方法^{*}

高 赞, 徐子钦, 王 涛, 周 浩

(云南大学信息学院 昆明 650504)

摘 要:为了解决多个尺度相关响应图最大值对应位置与目标真实位置存在偏差的问题,本文提出了一种采用多边形质心的相关滤波跟踪位置校正方法。首先,本文提出一种相关响应图评价指标对响应图的质量进行评价,并对当前帧跟踪结果进行可信判别;接着,对判别为不可信的跟踪结果采用多边形顶点的质心进行校正,以减少跟踪结果与目标正确位置的偏差。最后,在 OTB50、OTB-2015 和 UAV20L 三个基准视频集上对本文算法进行性能评估实验,本文算法在 OTB50、OTB-2015 和 UAV20L 的成功率曲线面积分别达到了 0.625、0.668 和 0.429,跟踪精度分别达到了 0.844、0.885 和 0.578。结果表明,与近年来主流的跟踪算法相比,本文算法在多种复杂场景下都取得了较优的成功率曲线面积和跟踪精度。

关键词: 目标跟踪;相关滤波;位置校正;多边形质心

中图分类号: TP391.4 TH701

文献标识码: A

国家标准学科分类代码: 510.4050

Polygon centroid based position correction method for correlation filter tracking

Gao Yun, Xu Ziqin, Wang Tao, Zhou Hao

(School of Information Science and Engineering, Yunnan University, Kunming 650504, China)

Abstract: To solve the problem of the deviation between the maximum position of the multi-scale correlation response map and the true position of the target, a polygon centroid based position correction method for correlation filter tracking is proposed. First, we proposed a correlation response map evaluation index to evaluate the quality of the response map, and make a credible judgment of the tracking result of the current frame. Then, we use the centroid of the polygon vertex to correct the tracking result that is judged to be unreliable to reduce the tracking result and target deviation from the correct position. Finally, the performance evaluation experiment of the algorithm in this paper is carried out on the three benchmark video sets of OTB50, OTB-2015 and UAV20L. The Area Under Curve of this algorithm in OTB50, OTB-2015 and UAV20L reached 0.625, 0.668 and 0.429 respectively, and the tracking accuracy reached 0.844, 0.885 and 0.578 respectively. The results show that compared with the mainstream tracking algorithms in recent years, our algorithm has achieved better success rate and tracking precision in a variety of complex scenarios.

Keywords: object tracking; correlation filter; position correction; polygon centroid

0 引 言

视频目标跟踪是计算机视觉领域的一个重要分支,由于该技术能够根据目标的初始信息确定后续视频帧中目标的位置、大小以及运动轨迹等信息,因此被广泛应用到人机交互、机器人视觉、导弹制导等各军事、民事领域中^[1]。尽管近年来目标跟踪算法在准确性和鲁棒性上有了突破性的进展,然而由于现实环境中存在相似物的干

扰、光照突变、部分或完全遮挡、运动模糊等复杂因素,使得目标跟踪仍然是一个具有挑战性的研究问题。

Bolme 等^[2]首次将相关滤波的概念引入到视频目标跟踪中,提出了最小输出误差平方和滤波(minimum output sum of squared error, MOSSE)算法,该算法能够高效的区分目标与背景,并且通过计算峰值旁瓣比以判断目标是否被遮挡,既保证了一定的跟踪准确性,同时也能够满足跟踪过程中实时性的要求。Wang 等^[3]提出的基于循环特征映射的边缘目标跟踪(Large Margin Object

Tracking with Circulant Feature Maps, LMCf)算法采用多峰检测的方法来判别目标是否被干扰,并提出了高置信度的模型更新策略。同样 Zhu 等^[4]认为由于跟踪场景中存在干扰,因此每一帧或固定间隔帧对模型进行更新是不合理的,因此他们的统一卷积网络(unified convolutional networks, UCT)算法提出了峰值噪声比(Peak-versus-noise Ratio, PNR),并根据 PNR 和最大响应值比值的历史均值的被动情况来判断模板是否更新。白中浩等^[5]基于多模型融合跟踪提出了一种可靠性判断依据以检测跟踪结果的可靠性,从而实现跟踪的稳定性。上述算法皆认为当目标受到多种复杂场景的干扰后,可能会使得模板学习到大量的错误信息而导致漂移,因此需要采用一个判别机制对当前帧的跟踪结果进行判别,从而提高算法的跟踪质量。

大多数基于相关滤波的目标跟踪算法是根据相关响应图最大值的对应位置作为目标的跟踪结果,然后用该跟踪结果作为新样本更新模板以适应目标的变化。Danelljan 等^[6]提出的基于精确尺度估计的相关滤波跟踪(discriminative scale space tracker, DSST)算法在二维的位置滤波器上采用了 33 个尺度以适应目标大小的变化,并通过调整跟踪框的比例来得到响应值最大的对应尺度和位置。基于 SAMF^[7]算法采用多尺度对目标候选区域计算的思想, Li 等^[8]提出了基于时空正则化的相关滤波(spatial-temporal regularized correlation filters, STRCF)算法,采用多个尺度对目标进行跟踪,并以相关响应图最大值的对应位置作为跟踪结果。Dai 等^[9]提出的基于自适应空间加权相关滤波(adaptive spatially-regularized correlation filters, ASRCF)算法采用两个滤波器分别估计目标的位置和尺度,不仅能够得到较精确的跟踪结果,并且也能简化运算的复杂度。翟敬梅等^[10]提出在预估位置范围内构建局部上下文外观模型,选择置信图最大值位置作为目标预测位置,从而完成目标的精确跟踪。可见,采用尺度滤波器有效促进了跟踪算法的尺度适应性。但是多个尺度中最大响应值并非总是与目标真实位置相符。

为了解决多个尺度最大响应值并非总是目标真实位置的问题,本文以 STRCF^[8]算法作为基本框架,提出一种采用多边形质心的相关滤波跟踪位置校正方法。本文的主要工作如下:1)提出相关响应图评价指标,并判别跟踪结果的优劣;2)采用多边形顶点的质心对不可信的跟踪结果进行校正;3)本文算法在目标跟踪基准(object tracking benchmark, OTB)50^[11]、OTB-2015^[12]和无人机(unmanned aerial vehicles, UAV)20L^[13]三个公开视频集上进行大量实验,在复杂场景下跟踪效果有较好的提升。

1 基础算法及改进框架

1.1 STRCF 算法原理

基于时空正则化的相关滤波(spatial-temporal regularized correlation filters, STRCF)^[8]跟踪算法通过将时间和空间正则化引入到相关滤波跟踪框架中,能够合理的近似于多训练样本中的基于空间正则化的相关滤波(spatially regularized correlation filter, SRDCF)^[14]形式,并且采用方向交替乘法(alternating direction method of multipliers, ADMM)有效求解,简化了 STRCF 运算的复杂度。与 SRDCF 相比,STRCF 使用当前帧的样本和上一帧学习到的滤波器 f_{t-1} 来训练该帧的滤波器 f_t , 因此 STRCF 不仅能够合理的处理边界效应,而且在外观发生变化和目标被遮挡的情况下具有更强的鲁棒性,同时不损失效率,在准确性和速度上都优于 SRDCF。为了简化符号,我们采用与 STRCF 中同样的表述方法,其目标函数为:

$$\arg \min_f \frac{1}{2} \left\| \sum_{d=1}^D x_t^d * f^d - y \right\|^2 + \frac{1}{2} \sum_{d=1}^D \|w \cdot f^d\|^2 + \frac{\mu}{2} \|f - f_{t-1}\|^2 \quad (1)$$

式中: $\sum_{d=1}^D \|w \cdot f^d\|^2$ 表示空间正则化, μ 表示正则化参数, $\|f - f_{t-1}\|^2$ 表示时间正则化。然后采用 ADMM 算法对式(1)进行优化,求解得到增广拉格朗日形式为:

$$L(w, g, s) = \frac{1}{2} \left\| \sum_{d=1}^D x_t^d * f^d - y \right\|^2 + \frac{1}{2} \sum_{d=1}^D \|w \cdot g^d\|^2 + \sum_{d=1}^D (f^d - g^d)^T s^d + \frac{\gamma}{2} \sum_{d=1}^D \|f^d - g^d\|^2 + \frac{\mu}{2} \|f - f_{t-1}\|^2 \quad (2)$$

式中:令 g 为引入的辅助变量且 $f = g$, γ 为步长参数。

然后采用 ADMM 算法交替求解 f, g, h 三个子问题,即:

$$\begin{cases} f^{(i+1)} = \arg \min_f \left\| \sum_{d=1}^D x_t^d * f^d - y \right\|^2 + \gamma \|f - g + h\|^2 + \mu \|f - f_{t-1}\|^2 \\ g^{(i+1)} = \arg \min_g \sum_{d=1}^D \|w \cdot g^d\|^2 + \gamma \|f - g + h\|^2 \\ h^{(i+1)} = h^{(i)} + f^{(i+1)} - g^{(i+1)} \end{cases} \quad (3)$$

式中: $^{(i+1)}$ 表示第 $i+1$ 次求解。将式(3)中的 f 子问题转换到傅里叶域上进行训练,即可得到最终的训练公式为:

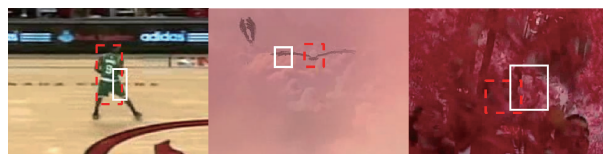
$$v_j(\hat{f}) = \frac{1}{\mu + \gamma} \left(I - \frac{v_j(\hat{x}_t) v_j(\hat{x}_t)^T}{\mu + \gamma + v_j(\hat{x}_t)^T v_j(\hat{x}_t)} \right) q, \quad (4)$$

式中: $\hat{f}$ 表示经过离散傅里叶变换的滤波器 f , 向量 $q = v_j(\hat{x}_i)\hat{y}_i + \gamma v_j(\hat{g}) - \gamma v_j(\hat{h}) + \mu v_j(\hat{f}_{i-1})$, $v_j(f) \in R^D$ 表示由 f 沿所有通道的第 j 个元素组成的向量。

1.2 算法分析

STRCF 在相关滤波跟踪框架中加入了时间和空间正则化系数, 不仅能够在遮挡和目标表观形变中具有更好的跟踪性能, 并且能够满足跟踪过程中的实时性要求。基于传统的相关滤波跟踪框架, STRCF 算法通常采用多个尺度相关响应图最大响应值的对应位置作为目标的跟踪结果, 在理论情况下是可行的。然而在实际的跟踪场景中, 目标往往会遇到背景混乱、光照变化、平面内/外旋转、遮挡、离开视野等场景的干扰, 此时对最大响应值对应的候选样本不加以判别直接将其作为跟踪结果可能会导致跟踪失败。

图 1 所示为 STRCF 算法在部分视频帧中的跟踪结果, 其中虚线框和实线框分别表示目标的正确位置和 STRCF 算法最大响应值对应的跟踪结果。从图中可以看出当目标受到外界或者自身的干扰时, 相关响应图最大值对应的位置或者尺度与目标正确位置逐渐产生了偏差。并且由于缺少对跟踪结果的可信判别, 使得算法在跟踪过程中逐帧累积的错误将最终导致跟踪失败。



(注: 虚线框和实线框分别表示目标正确位置和多个尺度最大响应值的对应位置)

图 1 算法在部分视频帧中的跟踪结果

Fig. 1 Tracking results of algorithm in some video frames

2 采用多边形质心的跟踪位置校正

2.1 改进算法总体框架

本文算法是基于 STRCF^[8] 算法的改进, 其总体框架如图 2 所示。在特征提取部分, 本文算法针对灰度视频序列融合方向梯度直方图 (histogram of oriented gradient, HOG)^[15]、灰度和强度通道 (intensity channels, IC)^[16] 特征进行特征提取; 而针对彩色视频序列, 本文算法则融合 HOG、灰度和颜色特征 (color name, CN)^[17] 特征完成对目标的表达。本文算法的改进点如图 2 阴影部分所示包括两个方面: 1) 计算相关响应图评价指标, 判别当前帧跟踪结果是否可信; 2) 针对不可信的跟踪结果, 采用多边形顶点的质心对当前帧跟踪结果进行校正。

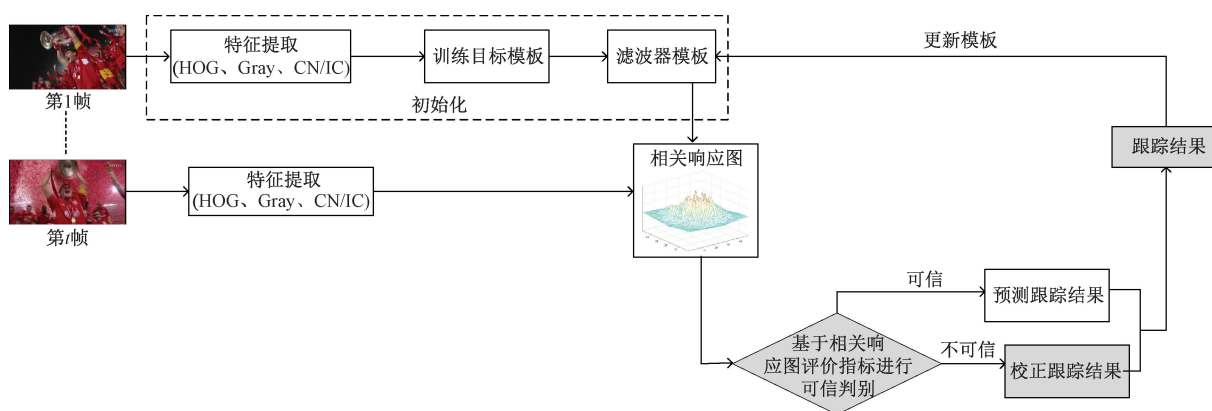


图 2 本文算法的总体框架

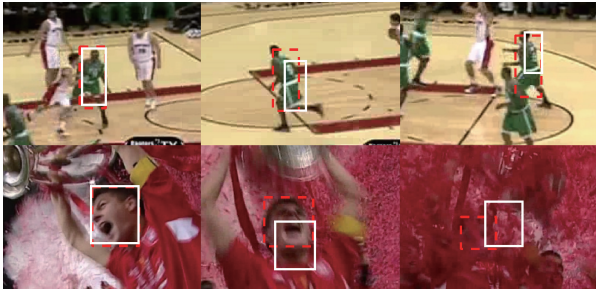
Fig. 2 Overall framework of our algorithm

2.2 跟踪结果的可信判别

理想情况下大多数相关滤波跟踪算法都以最大响应值的对应位置和尺度作为跟踪结果, 且每一帧跟踪结束之后都会采用当前帧跟踪结果作为新样本更新模板。当目标受到干扰时, 当前帧的最大响应值并不能完全作为评估跟踪结果的标准。如图 3 所示, 当场景中光照发生变化、混乱或者目标的外观、大小改变, 甚至受到遮挡等情况, 最大响应值容易忽略了跟踪过程中发生的波动, 此时相关响应图中主次峰之间的差异较小, 并且由于算法

缺少跟踪结果的可信判别机制, 因此跟踪算法无法确认该跟踪结果是否精确。然而不考虑该跟踪结果是否准确, 直接将其作为目标样本进行模板更新, 可能会引入大量的背景噪声或忽略目标信息, 从而使得目标模板发生漂移, 进而导致后续帧跟踪失败。

根据上述分析, 理想情况下, 响应矩阵是只有一个峰值的高斯分布图, 然而当目标受到各种复杂因素的干扰后, 其相关响应图会剧烈波动并呈现多峰状态, 统一卷积网络跟踪 (UCT)^[4] 算法中采用峰值噪声比 (PNR) 揭示



(注:虚线框和实线框分别表示目标的真实位置和多个尺度最大响应值的对应位置)

图 3 STRCF 在复杂场景中的跟踪结果

Fig. 3 STRCF tracking result in complex scenarios

相关响应图的分布情况。然而,由于每一帧响应图都具有不同的数值范围 $[R_{\min}, R_{\max}]$, 导致各视频帧之间或不同视频之间难以进行统一调参。因此,为了便于进行统一调参,本文对相关响应图做归一化处理,即将每一帧响应图的数值范围从 $[R_{\min}, R_{\max}]$ 线性拉伸到 $[0, 1]$, 采用了相关响应图评价指标 (RMEI) 对响应图的质量进行评价, RMEI 的计算方法为:

$$RMEI = \frac{R_{\max} - R_{\min}}{\frac{1}{n} \sum_{k=1}^n R_k - R_{\min}}, \quad (5)$$

式中: $R_{\max}$ 、 $R_{\min}$ 分别表示相关响应图的最大值和最小值, n 表示相关响应图中的响应值个数, R_k 表示相关响应图的第 k 个响应值。

根据式 (6) 可以判别当前帧跟踪结果是否可信。令 α, β 为阈值, D_t 表示第 t 帧跟踪结果的可信性, 则当前帧跟踪结果的判别方式为:

$$D_t = \begin{cases} 1, & \alpha < RMEI < \beta \\ 0, & \text{其他} \end{cases} \quad (6)$$

当 $D_t = 1$ 时, 表示当前帧采用最大响应值的对应位置作为跟踪结果较为可信, 能够作为新的样本更新目标模板; 当 $D_t = 0$ 时, 则表示当前的跟踪结果不可信, 此时目标可能处于复杂场景中或跟踪模板发生了漂移。

2.3 跟踪结果位置校正

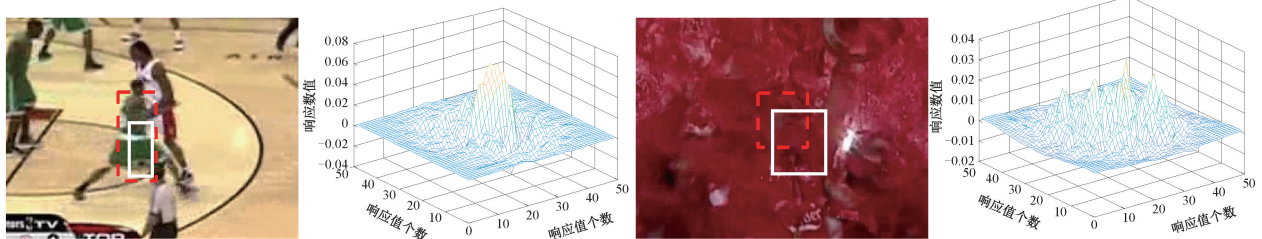
根据基于相关滤波目标跟踪算法的基本思想可知, 相关响应图中每个响应值都表示其所对应的样本与目标模板的匹配程度。对于单个响应图而言, 其响应值越大表示它所对应的样本与模板的匹配程度就越高, 因此我们总是期望找到响应值最大的对应位置作为目标的跟踪结果。然而对于多个响应图而言, 不同尺度之间的响应值并不具备绝对的可比性, 因此它们各自的最大响应值的对应位置都是可能的目标位置。基于相关滤波的跟踪算法采用多个尺度响应图中最大值对应位置作为目标的跟踪结果。但在实际的跟踪过程中, 目标通常会受到自身和外界的干扰, 多个尺度的最大响应值不一定总是对应目标的最佳位置, 如图 4 所示。相关响应图的主峰分散并出现多个峰值, 此时最大响应值不一定是目标的最佳位置。因此当目标受到现实环境的干扰时, 我们并不能因为某个尺度的响应值最大, 就盲目地采用该值对应位置作为当前帧跟踪结果。在本文算法中, 当采用式 (6) 判别当前帧跟踪结果不可信时, 采用多个尺度响应图各自的最大值来综合决定跟踪结果。当目标受到干扰时, 候选样本与目标之间的几种关系如表 1 所示。每个尺度的最大响应值都对应一个候选样本, 并且能够根据多个候选样本的中心点围成一个凸多边形或凹多边形, 当目标受到干扰时, 其正确位置的中心点大多数位于该多边形顶点的质心上。该多边形的每个顶点都对应着一个尺度的最大响应值, 每个顶点对应的最大响应值即该顶点的质量。每个顶点最大响应值对应的候选目标位置为 $L_i(x_i, y_i, s_i)$ 即该顶点的矢径。令 $L_\sigma(x, y, s)$ 表示多边形顶点的质心的矢径即跟踪结果的校正位置, 其计算方式为:

$$L_\sigma(x, y, s) = \sum_{i=1}^m R_{i, \max} L_i(x_i, y_i, s_i) / M, \quad (7)$$

M 表示多个顶点的总质量, 则:

$$M = \sum_{j=1}^m R_{j, \max} \quad (8)$$

其中, $R_{i, \max}$ 、 $R_{j, \max}$ 分别表示第 i, j 个尺度对应响应图的最大值; m 为尺度的个数; (x_i, y_i) 和 s_i 分别为第 i 个尺度的最大响应值对应候选目标的中心点位置和尺度。根据校正后的跟踪结果对目标模板进行更新。



(注:虚线框和实线框分别表示目标的真实位置和多尺度响应图最大值的对应位置)

图 4 干扰场景中目标的相关响应图

Fig. 4 Correlation response map of target in interference scene

表 1 目标正确位置与多个尺度跟踪框的关系

Table 1 The relationship between the correct position of the target and the multi-scale tracking boxes

	正确位置中心点位于多边形外	正确位置中心点位于多边形内
凸多边形		
凹多边形		

注:每组图第一幅点框表示目标真实位置,其余矩形框表示多个尺度对应位置;第二幅五边形表示多个尺度对应位置中心点围成的多边形,星号表示真实位置中心点;第三幅实线框和虚线框表示 STRCF 跟踪位置和修正位置。

3 实验结果与分析

3.1 实验环境及参数设置

本文算法的软件开发平台为 MATLAB R2017a。本文选取 OTB50^[11]、OTB-2015^[12]、UAV20L^[13] 三个基准视频集中的视频序列进行跟踪性能测试。OTB50 是 OTB-2015 中较难视频中组成的视频集,这两个基准视频集中含有 11 种具有挑战的场景,具体包括背景混乱 (background clutters, BC)、光照变化 (illumination variation, IV)、遮挡 (occlusion, OCC)、离开视野 (out-of-view, OV)、快速运动 (fast motion, FM)、运动模糊 (motion blur, MB)、平面内旋转 (in-plane rotation, IPR)、平面外旋转 (out-of-plane rotation, OPR)、尺度变化 (scale variation, SV)、低分辨率 (low resolution, LR)、目标表观形变 (deformation, DEF)^[12], 其中每个视频序列至少包含了一种挑战属性。UAV20 L 则是 UAV 视频集中较长视频序列组成的一个子集,其跟踪难度较大,并且也包含了 12 个挑战属性,分别是 SV、纵横比变化 (aspect ratio-change, ARC)、LR、FM、完全遮挡 (full occlusion, FOCC)、部分遮挡 (partial occlusion, POCC)、OV、BC、IV、视角变化 (viewpoint change, VC)、相机运动 (CM)、相似目标 (similar object, SO)^[13], 与 OTB 视频集相同,UAV20 L 中每个视频也包含了两种或两种以上的挑战属性。

由于本文提出的 RMEI 能够对不同的视频序列设置统一的阈值,因此本文设定阈值 α, β , 以判别响应图的质量及确定跟踪结果的优劣。本文算法在 $0 < 3 - \alpha < 0.3$ 、 $\beta = 4$ 时可以取得较好的跟踪效果。本实验设置 $\alpha = 2.89$ 、 $\beta = 4$, 其余参数设置与 STRCF 一致。

3.2 评估指标

实验中采用 OTB 视频集,选择一次性通过评价 (OPE)^[12] 中的成功率曲线面积 (area under curve, AUC)^[12]、精度 (precision)^[18] 和中心像素点距离误差 (center location error, CLE) 三个指标来评估本文算法的跟踪效果。令跟踪结果的矩形框为 A_t 和目标真实位置标注框为 A_{gt} , 则重叠得分被定义为:

$$S = |A_t \cap A_{gt}| / |A_t \cup A_{gt}| \quad (9)$$

其中, $|\cdot|$ 表示区域的像素数目。当某一帧的重叠得分大于所设定的阈值,则认为该帧成功,总的成功帧数占所有帧的百分比即为成功率,根据该成功率绘制一条曲线,则曲线下的面积即为成功率曲线面积;而精度则是计算每一帧的像素误差,即计算跟踪结果矩形框中心点与目标真实位置中心点之间的距离,该像素误差小于给定阈值的帧数占总帧数的比例为该帧对应的精度。CLE 是跟踪结果与目标真实位置之间的欧式距离,其值越小表示跟踪精度越高。

3.3 定量分析

本文在 OTB50^[11]、OTB-2015^[12] 和 UAV20L^[13] 三个基准视频集中分别对算法进行性能测试,得到各自的 AUC 曲线和精度曲线。为了更好地评价本文算法的跟踪性能,本文将在 OTB50 和 OTB-2015 两个基准视频集中与近年来主流的跟踪算法:无监督深度跟踪算法 (unsupervised deep tracking, UDT)^[19]、大位移跟踪与相似估计算法 (large-displacement tracking via estimation of similarity, LDES)^[20]、多线索相关滤波跟踪算法 (multi-cue correlation filters trackers, MCCT-H)^[21]、STRCF^[8]、有效卷积算子 (efficient convolution operators, ECO-HC)^[22]、背景感知相关滤波算法 (background-aware correlation filters, BACF)^[23]、模板和



像素学习器集成算法 (sum of template and pixel-wise learners, Staple)^[24]、核相关滤波算法 (kernel correlation filters, KCF)^[25]、自动时空正则化补充跟踪算法 (tracker with automatic spatio-temporal regularization supplementray material, AutoTrack)^[26] 进行对比。在 UAV20 L 视频集中本文算法将与 STRCF、MCCT-H、ECO-HC、异常抑制相关滤波算法 (aberrance repressed correlation filters, ARCF)^[27]、AutoTrack^[26]、BACF 六种算法进行对比。其中 ARCF、UDT 和 LDES 算法是 2019 年主流的跟踪算法,而 AutoTrack 是 2020 年的主流跟踪算法,并且 ARCF 和 AutoTrack 针对无人机进行跟踪的算法。

为了更直观的评价本文算法的跟踪效果,图 5 中展示了 10 种跟踪算法在 OTB-2015 和 OTB50 两个视频集上的 AUC 图和跟踪精度图,从图中数据分析可知,本文

算法的 AUC 和跟踪精度在两个视频集中均达到了最高。在 OTB-2015 基准视频集中,本文算法的 AUC 达到了 66.8%,相比原基础算法提高了 1.4%,与 ECO-HC 相比提高了 2.9%,与 2020 年主流跟踪算法 AutoTrack 相比提高 8.1%;在跟踪精度方面,本文算法的精度也达到了 88.5%,相比原基础算法提高了 2.1%,与 ECO-HC 相比也提高了 2.9%,与 AutoTrack 算法相比提高了 9.6%。在 OTB50 基准视频集中,本文算法的 AUC 达到 62.5%,相比 STRCF 提高了 2%,相比 ECO-HC 提高了 3.3%,与 AutoTrack 相比提高了 9.7%。并且本文算法在跟踪精度上也达到了 84.4%,相比 STRCF 提高了 1.8%,相比 ECO-HC 提高了 3%,与 AutoTrack 相比跟踪精度提高了 11.2%。可以看出本文算法不管是在 AUC 还是跟踪精度上都取得了最优的效果。

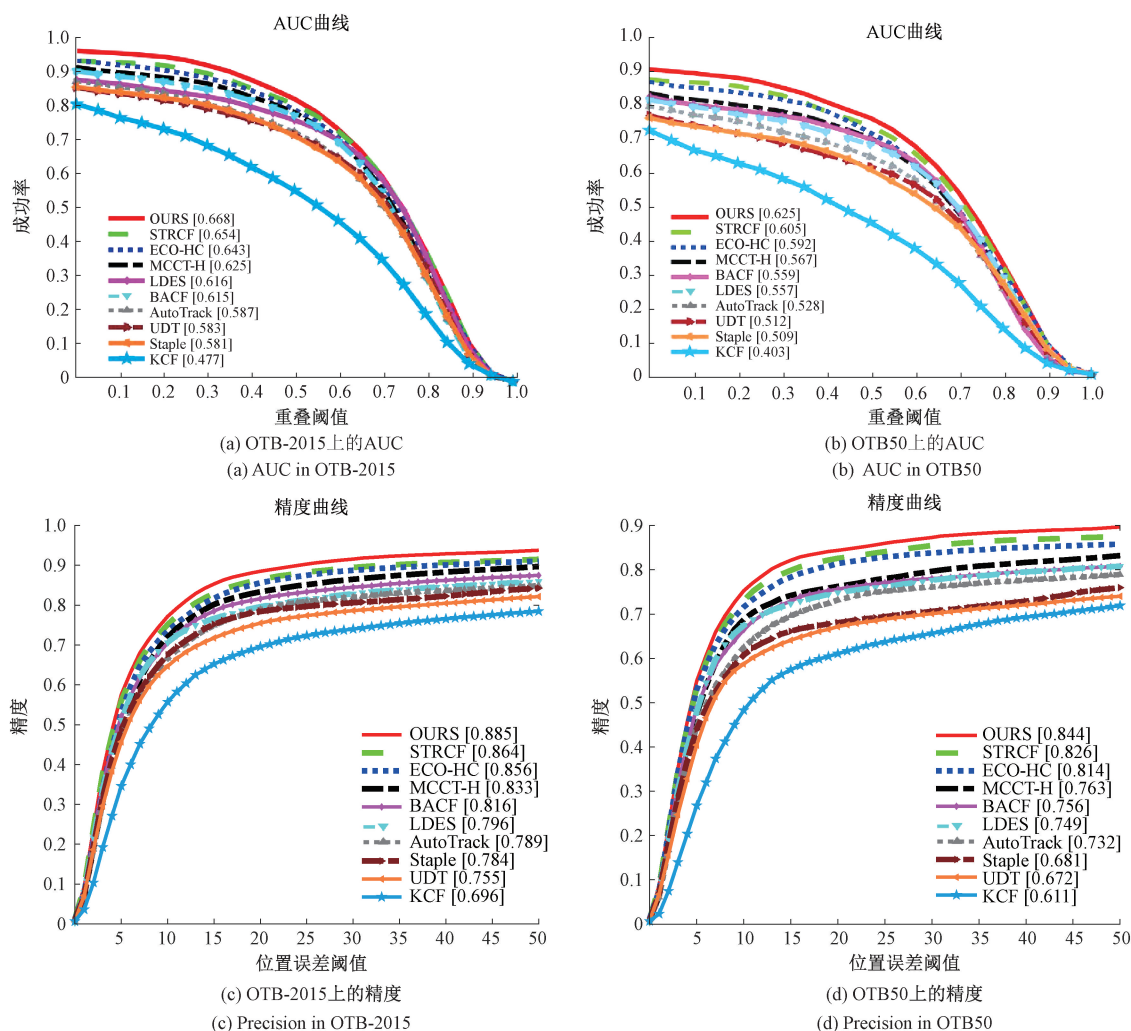


图 5 9 种算法在 OTB-2015 和 OTB50 上的 AUC 和精度

Fig. 5 AUC and precision of nine algorithms on OTB-2015 and OTB50 databases

图 6 中展示了本文算法 UAV20 L 视频集中的 AUC 和跟踪精度。从图中可以看出本文算法在 AUC 和跟踪

精度中都能排在首位,其中本文算法的 AUC 达到了 0.429,相比原基础算法 STRCF 提高了 1.9%。而跟踪精

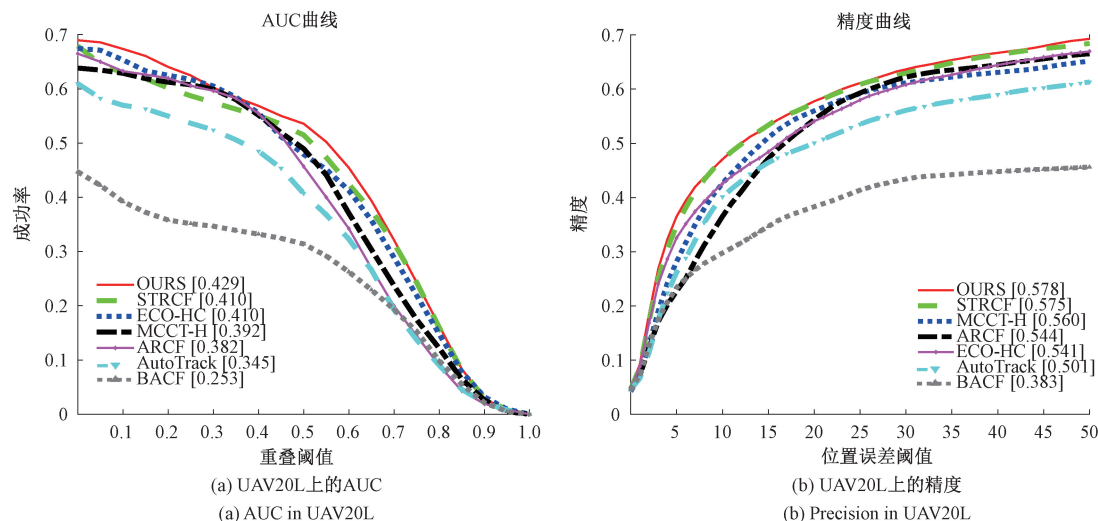


图6 7种算法在UAV20 L上的AUC和跟踪精度

Fig. 6 AUC and precision of seven algorithms on UAV20 L databases

度达到了0.578,与原基础算法相比提高了0.3%。并且与近两年在UAV视频集中主流的ARCF和AutoTrack相比,本文算法的AUC分别提高了4.7%和8.4%,而跟踪精度则分别提高了3.4%和7.7%。

为了进一步分析本文算法在不同场景下的跟踪性能,本文在OTB-2015、OTB50和UAV20 L三个基准视频集上进行了实验,由于OTB-2015是在OTB50基础上扩展的视频集,因此这里仅展示本文算法在OTB-2015中各种挑战场景下的性能。结果如图7~9所示。从图7中可以清晰直观的看出,在OTB-2015视频集中,本文算法的AUC和精度始终处于最优或次优水平。在AUC方面,本文算法在9种属性下均取得最优,而在低分辨率和尺度变换两种属性中也仅以微弱的差距排在第2;在精度方面,本文算法在8种属性下取得最优,而在快速运

动、低分辨率和离开视野场景中仅次于ECO-HC排在第2名。本文算法是在目标受到场景干扰后,对不可信的跟踪结果采用多边形顶点的质心进行校正,因此在AUC和跟踪精度上都了一定的提高,并且对大多数场景都有着较好的稳定性,特别是在背景混乱、光照变化和遮挡场景中,本文算法的AUC分别达到了68.9%、67.4%和65.1%,相比排在第2名的算法提高了3.7%、2.1%和2.2%。而跟踪精度在背景混乱、平面内旋转和光照变化表现较为优越,分别达到了91.3%、83.8%和86.8%,相比排在第2名的STRCF算法提高了3.7%、2.3%和2.7%。图8~9中展示了本文算法与其他6种算法在UAV20 L视频集中部分挑战属性的AUC和跟踪精度,从图中可以看出,本文算法在相机运动、纵横比变化、尺度变化中的AUC分别为0.412、0.356和0.413,相比原基

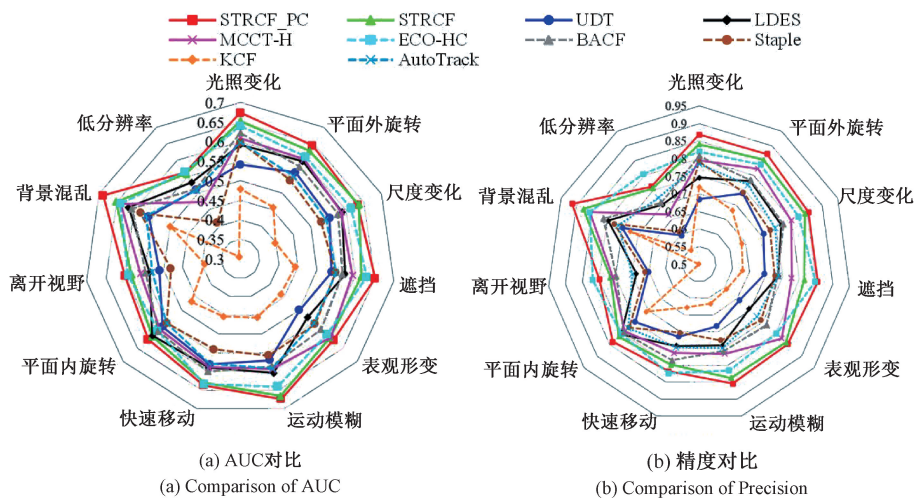


图7 在OTB-2015下11种不同属性的AUC和精度对比

Fig. 7 Comparison of the AUC and precision of 11 different attributes on OTB-2015 database

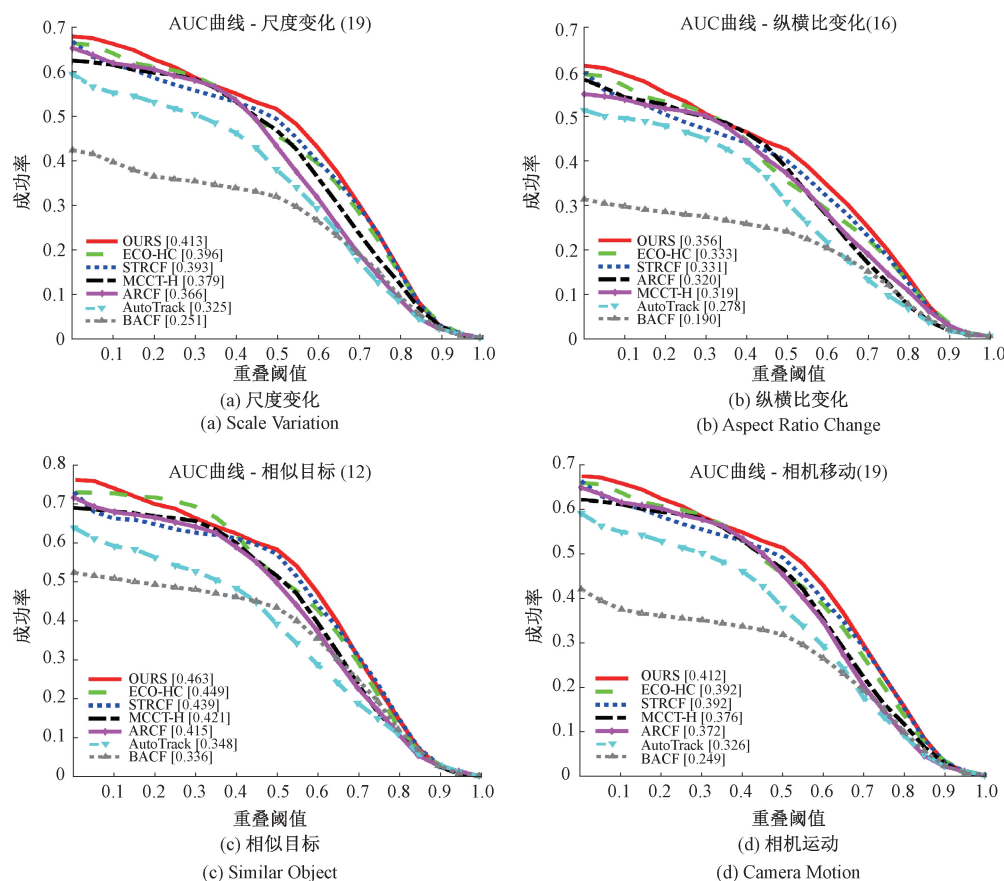


图 8 UAV20L 中部分属性的 AUC 曲线

Fig. 8 AUC plots under different attributes of UAV20L

础算法分别提高了 2%、2.3% 和 2%。而在相似目标、部分遮挡、相机运动中的跟踪精度分别为 0.567、0.568 和 0.555,与排在第 2 位的算法相比分别提高了 1.7%、0.4% 和 0.2%。

3.4 定性分析

为了更直观地评估本文算法在不同场景中的跟踪性能,本文选取了 8 个具有代表性的视频序列,对算法在背景混乱、表观形变、遮挡、平面内旋转、平面外旋转等特定场景下分别进行定性分析,结果如图 10~13 所示。

1) 背景混乱

图 10 所示为 Basketball、Soccer 视频中部分视频帧的跟踪结果和 CLE 图,对于 Basketball 视频,由于目标运动员在移动过程中产生形变和旋转,并且周围一致存在相似运动员的干扰,从第 189~649 帧 STRCF 算法虽然能跟踪到目标,但其鲁棒性较低,而 BACF 则偏离目标。第 650 帧光照突然增强且目标受到了其他干扰物的遮挡,此时 Staple 算法偏移至该干扰物上。第 660 帧场景中的光照恢复正常,但 ECO-HC 和 STRCF 算法仍然有一定程度的偏离。而在整个跟踪过程中本文算法能够维持正确的跟踪。对于 Soccer 视频,其跟踪难度较大,目标存在被遮挡、干扰、快速运动以及运动模糊等多种因素的干扰。

在开始阶段所有跟踪算法均能准确跟踪到目标。从第 58~76 帧目标周围存在红色干扰物,且目标本身由于跳跃而造成场景模糊,此时 STRCF、UDT、KCF、LDES 均有不同程度的偏离。第 97 帧开始目标逐渐被奖杯遮挡,并且此时红色丝带的干扰十分严重,导致 ECO-HC、BACF、KCF、MCCT-H 逐渐开始偏离目标。第 133 帧开始仅有本文算法和 KCF 能够跟上目标。而在整个跟踪过程中仅有本文算法能够维持正确的跟踪。并且从这两个视频序列的 CLE 图中可知,由于目标所在场景较为混乱,存在相似物干扰时,本文算法能够及时找到错误的候选样本并及时修正,并且采用多特征融合的方式也能够减少其他因素对目标的干扰,使得与其他算法相比,本文算法的 CLE 曲线更贴近水平轴,维持了良好的稳定性。

2) 遮挡

图 11 所示为 Girl、Girl2 视频中部分视频帧中的跟踪结果和 CLE 图,这两个视频序列包括了遮挡、尺度变化和旋转。对于 Girl 视频,在开始阶段目标周围没有任何干扰,此时所有算法均能准确跟踪到目标。第 85~115 帧和第 169~242 帧,目标都发生旋转,MCCT-H、Staple、ECO-HC、AutoTrack 算法均有不同程度的偏离。第 429~469 帧由于目标被遮挡,此时 KCF、MCCT-H、Staple 和 BACF 算

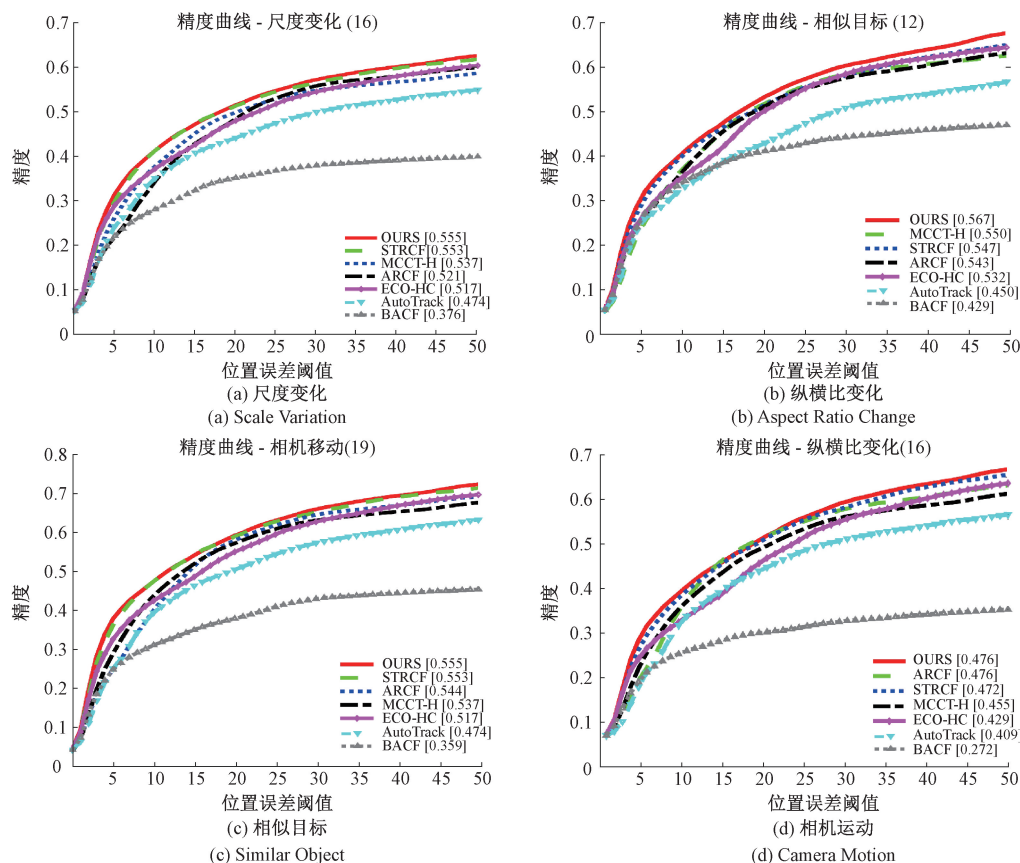


图9 UAV20L 中部分属性的精度曲线

Fig.9 Precision plots under different attributes of UAV20L

法均偏移至遮挡物上,而 AutoTrack 算法则偏移至目标的头发上。在这整个跟踪过程中本文算法能够保持正确的跟踪。对于 Girl2 视频,第 110 帧目标受到了干扰物的完全遮挡,此时所有算法均无法跟踪到目标。第 289~351 帧目标重新出现一段时间后,本文算法、STRCF 和 MCCT-H 皆能跟踪到目标,而 UDT、BACF、ECO-HC、LDES、AutoTrack、Staple 和 KCF 算法均偏离目标,始终漂移在背景中。并且从 CLE 图中我们可以看到,本文算法的 CLE 曲线与其他算法相比更加贴近水平轴,即本文算法的跟踪结果与目标真实位置的像素误差更小。当目标被局部或完全遮挡时,其相关响应图最大值的对应位置可能是遮挡物的位置,而非目标的真实位置,因此本文算法采用多边形顶点的质心校正不可信的跟踪结果,使得算法能够在目标重新出现后及时找到,有效提高跟踪算法的鲁棒性。

3) 表观形变

图 12 所出为 Skater、Bird1 视频中部分视频帧中的跟踪结果和 CLE 图,这两个视频序列包括了目标形变。对于 Skater 视频,目标开始运动后其外观发生形变,第 16 帧 LDES 算法逐渐发生偏离。第 47~81 帧目标在运动过程中发生旋转并且其外观也在不断发生变化,此时 ECO-HC、MCCT-H、Staple 和 LDES 算法发生了不同程度的偏

离,而在整个跟踪过程中本文算法能够正确持续的跟踪。对于 Bird1 视频,目标周围存在许多相似的干扰物,并且目标在飞行过程中外观不断改变,加之有很长的一段时间内目标消失在视野内,因此其跟踪难度较大。第 18~51 帧目标在飞行过程中不断扇翅膀,其外观发生变化,此时 ECO-HC、KCF、LDES 和 UDT 均有发生了漂移。第 153 帧目标被云层完全遮挡消失在视野内,此时所有跟踪算法都无法找到目标。第 182~261 帧当目标重新出现在视野内,且其外观也在变化的情况下,仅有本文算法能够较为准确的跟踪到目标,而其他算法则都偏移至背景或其他相似物上。从 CLE 图中也可看出,虽然本文算法的 CLE 值在这两个视频均排在第 2 位,但与其他算法相比本文算法的 CLE 曲线仍然更贴近水平轴,能够得到较好的跟踪结果。因此当目标发生形变时,本文算法使用多边形顶点的质心的矢径校正跟踪位置的方法能够减缓形变对目标模板带来的干扰,并且与其他 7 种算法相比,本文算法在鲁棒性和准确性上仍然有很大的提升。

4) 平面内/外旋转

图 13 所示为 Sylvester、Panda 视频中部分视频帧的跟踪结果和 CLE 图,这两个视频序列包括背景混乱和平面向外旋转。对于 Sylvester 视频,在开始阶段所有跟踪

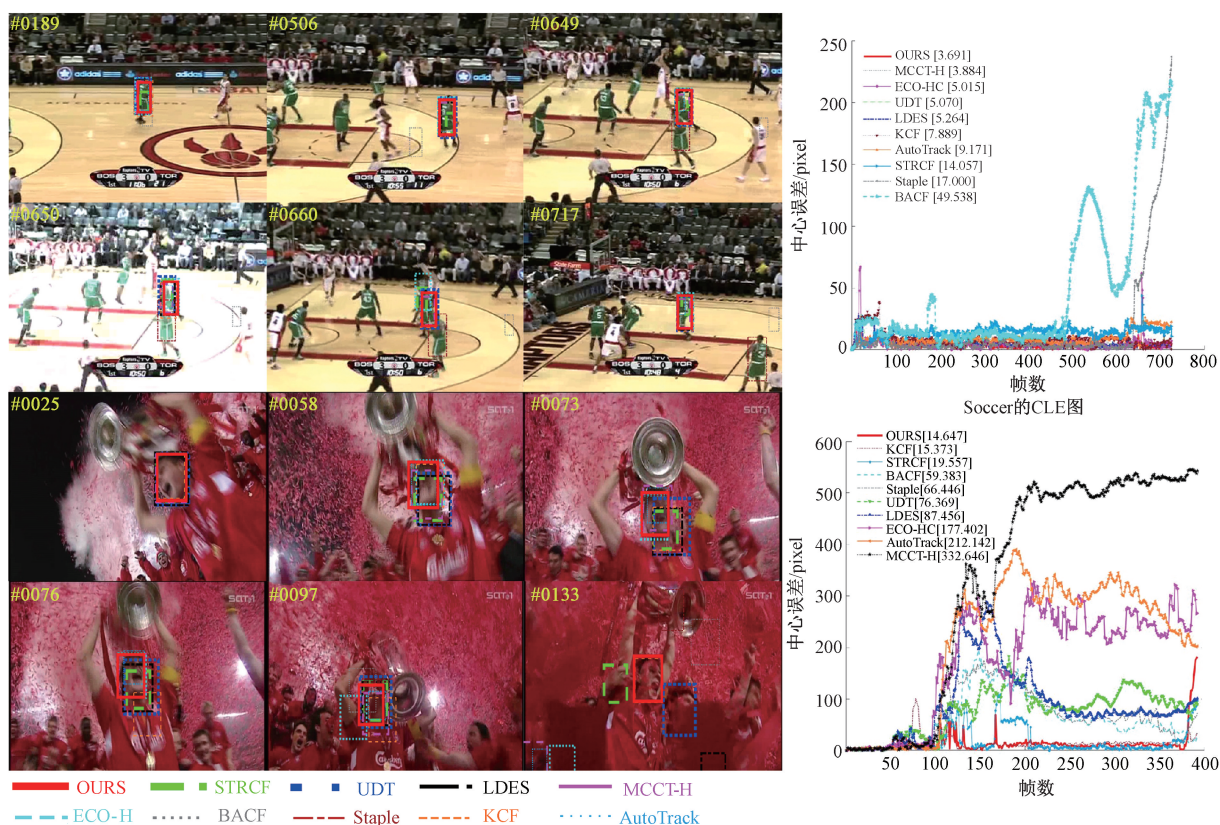


图 10 Basketball、Soccer 中部分视频帧的跟踪结果和 CLE 图

Fig. 10 Tracking results and CLE plots of some video frames in Basketball and Soccer

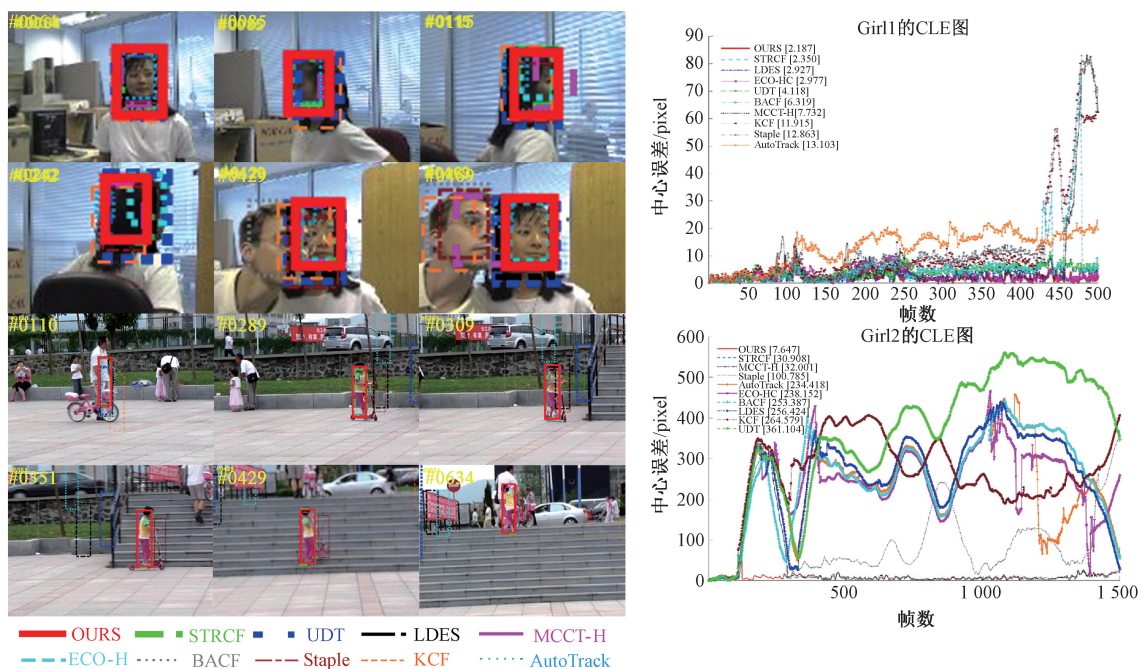
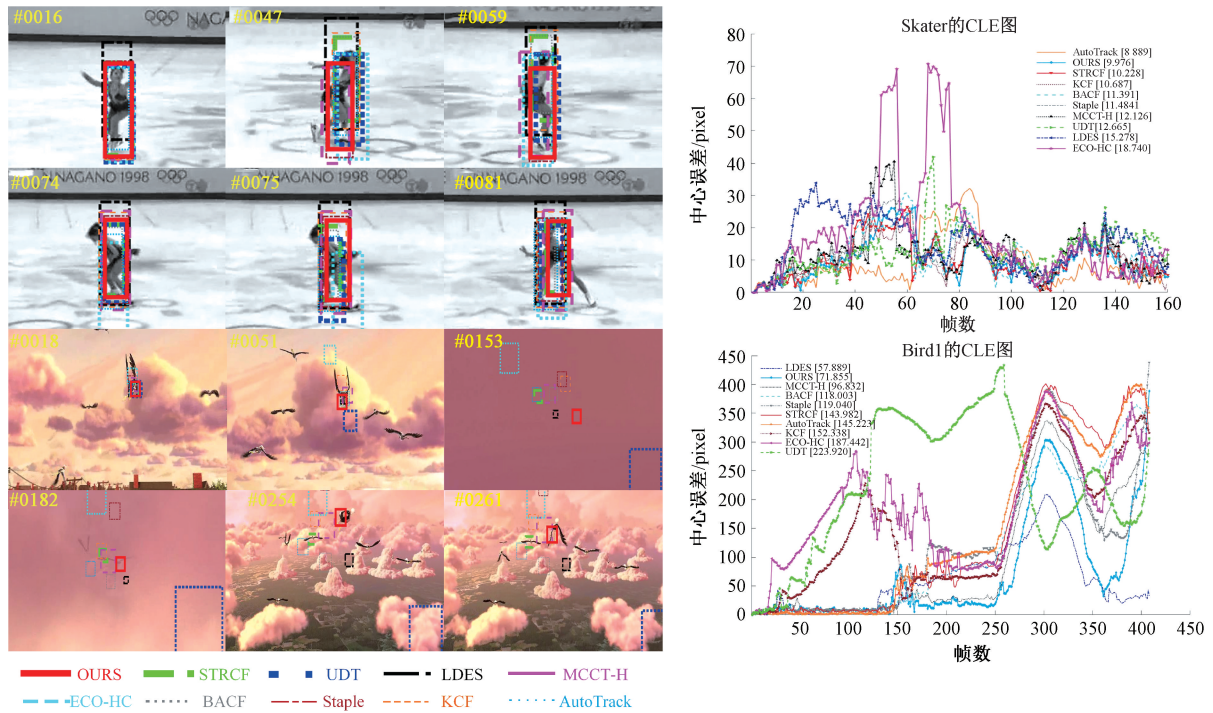
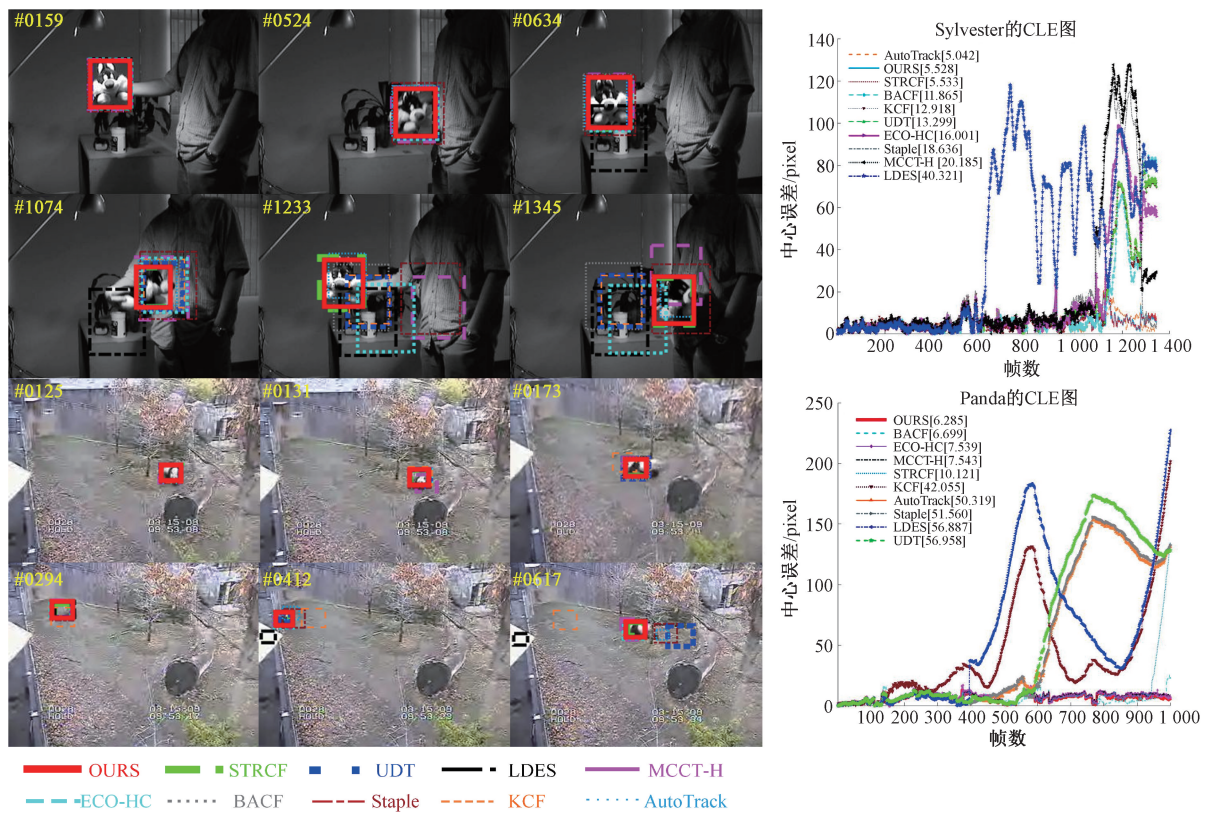


图 11 Girl、Girl2 中部分视频帧的跟踪结果和 CLE 图

Fig. 11 Tracking results and CLE plots of some video frames in Girl and Girl2

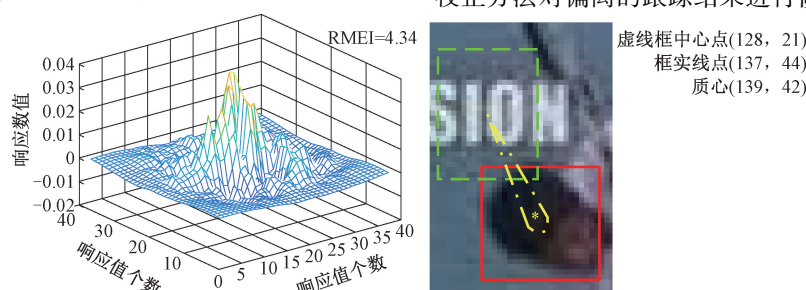
图 12 *Skater*、*Bird1* 中部分视频帧的跟踪结果和 CLE 图Fig. 12 Tracking results and CLE plots of some video frames in *Skater* and *Bird1*图 13 *Sylvester*、*Panda* 中部分视频帧的跟踪结果和 CLE 图Fig. 13 Tracking results and CLE plots of some video frames in *Sylvester* and *Panda*

算法均能准确跟踪到目标。第 524 帧目标在光照和阴影处移动,并且此时目标有轻微旋转的动作,此时 Staple 算法虽然能跟踪到目标,但其鲁棒性在逐步降低。第 634 帧目标开始旋转,此时 LDES 算法逐渐开始偏离目标。第 1 074~1 345 帧随着目标所在场景中光线不断变化以及目标在不断旋转,此时除了本文算法、AutoTrack 和 STRCF 能够较为准确跟踪到目标之外,其他算法都偏移至背景中。对于 Panda 视频,在开始阶段所有算法均能准确跟踪到目标。第 131~294 帧,目标发生了缓慢的旋转,此时 KCF 逐渐偏离目标。第 412~617 帧由于目标在发生旋转的同时还伴随着尺度的变化,因此 LDES、Staple、KCF、UDT、AutoTrack 均无法跟踪到目标,而此时仅有本文算法、STRCF、MCCT-H、ECO-H、BACF 能够较为准确的跟踪目标。从 CLE 图中可知,其值越小表示算法的跟踪结果越精确,虽然本文算法的 CLE 曲线在

Sylvester 视频中次于 AutoTrack,但与其他算法相比仍然更加贴近水平轴。因此本文算法与其他算法相比具有较为精确的跟踪结果。当目标发生旋转时,由于本文算法采用多边形顶点的质心的校正不可信的跟踪结果,使得算法能够及时跟上目标发生的旋转。

3.5 案例分析

图 14 所示为当目标受到干扰时,最大响应值的跟踪结果和多边形质心的位置,该视频帧来自 OTB-2015 视频集 KiteSurf 视频。从图中可以看出,当目标受到遮挡时,相关响应图呈现多个峰值,相关响应图最大值对应位置 (128, 21) 远远偏离了目标真实位置 (137, 44),无法再作为跟踪结果继续完成对目标的跟踪。而此时多个尺度响应图最大值形成的多边形顶点的质心 (139, 42) 更逼近目标真实位置的质心。由此可见,采用多边形质心的位置校正方法对偏离的跟踪结果进行修正可行的。



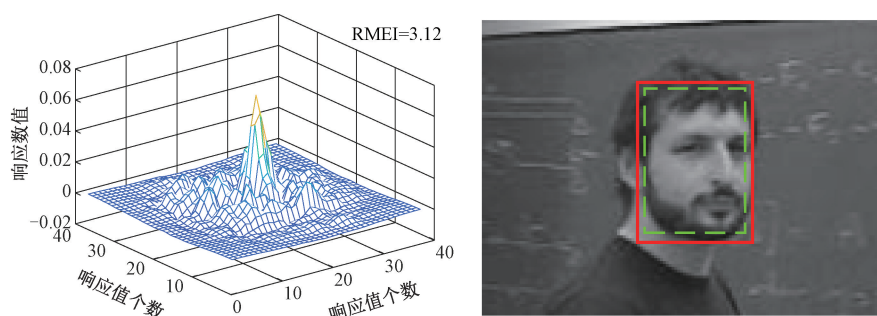
(其中虚线框和实线框分别表示最大响应值对应的位置和目标的真实位置,点线五边形表示由多个尺度响应图形成的多边形,“*”则表示该多边形顶点质心的位置)

图 14 目标受到干扰时的相关响应图和跟踪结果

Fig. 14 Correlation response map and tracking result when the target is interfered

值得说明的是,理想情况下或简单场景下相关滤波跟踪方法的相关响应图大多会呈现单尖峰状,其值对应位置更加逼近目标的真实位置,此时最大响应值对应的位置是较为可靠的跟踪结果,无需对该跟踪结果进行修正。图 15 所示为本文算法在简单场景下的相关响应图以及对应的跟踪结果示例,该视频帧来自 OTB-2015 视频

集 Freeman1 视频。从图中可以看出,目标背景较为干净,相关响应图最大峰值较为尖锐,且与其他峰值差异较大,最大值对应位置更为接近目标真实位置。而最右侧偏离目标的尺度框将导致质心偏离目标,若此时采用多边形质心的位置校正方法对该跟踪结果进行修正,反而会使修正后的跟踪结果与目标真实位置产生偏移。因



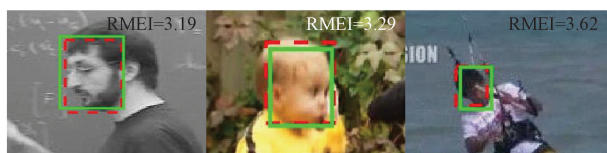
(其中虚线框和实线框分别表示最大响应值对应的位置和目标的真实位置)

图 15 目标不受到干扰时的相关响应图和跟踪结果

Fig. 15 Correlation response map and tracking results when the target is not interfered

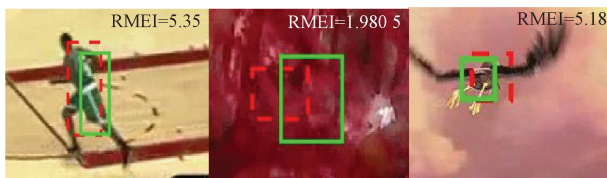
此,只有在目标受到如背景混乱、遮挡、光照变化等复杂场景因素干扰且导致最大响应值对应位置偏离真实位置时,本文算法才采用多边形质心的位置校正方法对当前帧跟踪结果进行修正。

本文算法采用 RMEI 指标对最大响应值对应位置是否偏离真实位置进行判别。由于 RMEI 指标计算过程中对响应值进行了归一化处理,很容易发现该指标与跟踪结果的可靠性之间呈现出一定的关联。图 16 所示为一些可信和不可信跟踪结果对应的 RMEI 值。在简单场景下最大响应值对应位置较为可靠时,RMEI 通常处于中间值区域(如图 16(a)所示),本实验设置为区间 $[2.89, 4]$;在复杂场景下最大响应值偏离真实位置时,RMEI 通常较小或者较大(如图 16(b)所示)。定量分析结果也表明,本文算法采用的 RMEI 判别指标和多边形质心修正方法在一定程度上促进了跟踪的成功率和精度。



(a) 可信跟踪结果对应的 RMEI 值

(a) RMEI values of some trusted tracking results



(b) 不可信跟踪结果对应的 RMEI 值

(b) RMEI values of some untrusted tracking results

(其中实线框和虚线框分别表示最大响应值对应的位置和目标的真实位置)

图 16 可信和不可信跟踪结果对应的 RMEI 值

Fig. 16 RMEI values of some trusted and untrusted tracking results

4 结 论

本文在时空正则化相关滤波跟踪算法基础上,提出了一种采用多边形质心的相关滤波跟踪位置校正方法。由于目标在受到各种复杂因素的干扰时,其最大响应值对应的位置并非总是逼近目标的真实位置,因此首先计算相关响应图评价指标以评价响应图的质量,进而判断当前帧跟踪结果是否可信;最后对不可信的跟踪结果,基于多边形顶点的质心对当前帧跟踪结果的位置和尺度进行校正。与近几年主流跟踪算法的对比实验表明,本文算法取得了较优的综合 AUC 和精度,在实际复杂场景下尤其是背景混乱、遮挡、光照变化、旋转、纵横比变化等场

景下具备了更好的鲁棒性和准确性。在未来工作中跟踪结果的可信判别和位置校正方法值得进一步研究。

参考文献

- [1] 孟球, 杨旭. 目标跟踪算法综述[J]. 自动化学报, 2019, 45(7): 1244-1260.
MENG L, YANG X. A survey of object tracking algorithms[J]. Acta Automatica Sinica, 2019, 45(7): 1244-1260.
- [2] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010: 2544-2550.
- [3] 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, 2017: 4021-4029.
- [4] ZHU Z, HUANG G, ZOU W, et al. Uct: Learning unified convolutional networks for real-time visual tracking[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2017: 1973-1982.
- [5] 白中浩, 朱磊, 李智强. 多模型融合和重新检测的高精度鲁棒目标跟踪[J]. 仪器仪表学报, 2019, 40(9): 135-144.
BAI ZH H, ZHU L, LI ZH Q. High accuracy and robust object tracking based on multi-model fusion and re-detection [J]. Chinese Journal of Scientific Instrument, 2019, 40(9): 135-144.
- [6] DANELLJAN M, HÄGER G, KHAN F S, et al. Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(8): 1561-1575.
- [7] LI Y, ZHU J. A scale adaptive kernel correlation filter tracker with feature integration[C]. European Conference on Computer Vision, Springer, Cham, 2014: 254-265.
- [8] 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, 2018: 4904-4913.
- [9] DAI K, WANG D, LU H, et al. Visual tracking via adaptive spatially-regularized correlation filters [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 4670-4679.
- [10] 翟敬梅, 刘坤. 特征点辅助的时空上下文目标跟踪与定位[J]. 仪器仪表学报, 2017, 38(11): 2839-2848.
ZHAI J M, LIU K. Object tracking and location with spatio temporal context assisted by key points [J].



- Chinese Journal of Scientific Instrument, 2017, 38(11): 2839-2848.
- [11] WU Y, LIM J, YANG M H. Online object tracking: A benchmark[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2013: 2411-2418.
- [12] WU Y, LIM J, YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1934-1848.
- [13] MUELLER M, SMITH N, GHANEM B. A benchmark and simulator for uav tracking[C]. European Conference on Computer Vision, Springer, Cham, 2016: 445-461.
- [14] 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 (ICCV), IEEE, 2015: 4310-4318.
- [15] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), IEEE, 2005, 1: 886-893.
- [16] FELSBERG M. Enhanced distribution field tracking using channel representations[C]. Proceedings of the IEEE International Conference on Computer Vision Workshops, 2013: 121-128.
- [17] 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, IEEE, 2014: 1090-1097.
- [18] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (voc) challenge[J]. International Journal of Computer Vision, 2010, 88(2): 303-338.
- [19] WANG N, SONG Y, MA C, et al. Unsupervised deep tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 1308-1317.
- [20] LI Y, ZHU J, HOI S C H, et al. Robust estimation of similarity transformation for visual object tracking[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 8666-8673.
- [21] WANG N, ZHOU W, TIAN Q, et al. Multi-cue correlation filters for robust visual tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4844-4853.
- [22] DANELLJAN M, BHAT G, SHAHBAZ K F, et al. Eco: Efficient convolution operators for tracking[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2017: 6638-6646.
- [23] KIANI G H, FAGG A, LUCEY S. Learning background-aware correlation filters for visual tracking[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1135-1143.
- [24] 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, 2016: 1401-1409.
- [25] 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, 2014, 37(3): 583-596.
- [26] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11923-11932.
- [27] HUANG Z, FU C, LI Y, et al. Learning aberrance repressed correlation filters for real-time uav tracking[C]. Proceedings of the IEEE International Conference on Computer Vision, 2019: 2891-2900.

作者简介



高赞(通信作者), 2001 年于西安电子科技大学获得学士学位, 2007 年于西安电子科技大学获得硕士学位, 2014 年于云南大学获得博士学位。现为云南大学副教授, 主要研究方向为视频运动目标跟踪。

E-mail: gaugegao@163.com

Gao Yun (Corresponding author) received her B. Sc. degree from Xidian University in 2001, received her M. Sc. degree from Xidian University in 2007, and received her Ph. D. degree from Yunnan University in 2014. Now she is an associate professor of Yunnan University. Her main research interest is visual moving object tracking.



徐子钦, 2018 年于云南财经大学获得学士学位。现为云南大学硕士研究生, 主要研究方向为视频运动目标跟踪。

E-mail: carnelianxu@163.com

Xu Ziqin received her B. Sc. degree from Yunnan University of Finance and Economics in 2018. Now she is a M. Sc. candidate in Yunnan University. The main research interest is visual moving object tracking.