

基于自适应特征融合与上下文感知的目标跟踪

纪元法^{1,2}, 何传骥^{1,2}, 孙希延^{1,2*}, 郭宁^{1,2}

¹桂林电子科技大学广西精密导航技术与应用重点实验室, 广西 桂林 541004;

²卫星导航定位与位置服务国家地方联合工程研究中心, 广西 桂林 541004

摘要 目标外观的描述对相关滤波跟踪器的性能有很大的影响,单一特征难以准确描述目标外观,基于多特征的目标外观描述在复杂场景下有着更加优异的性能。为了提高复杂场景下目标跟踪的鲁棒性,提出一种基于多特征自适应融合与上下文感知的目标跟踪算法。首先引入上下文感知框架,提取目标周围 4 个上下文图像块的一层卷积特征作为背景信息。由于单一特征难以准确描述目标外观,采用两个相关滤波器来提取多种特征。第一个滤波器通过卷积神经网络提取三层卷积特征作为深度特征,第二个滤波器提取方向梯度直方图和颜色直方图信息作为浅层特征,然后自适应融合深、浅特征。最后,利用平均峰值相关能量来评估响应的置信度,并决定是否对模型进行更新。所提算法在 OTB-2013 数据集上进行测试,结果表明,该算法在准确率和成功率方面都有很好的表现,与其他优秀的跟踪算法相比,具有更好的跟踪性能。

关键词 图像处理; 机器视觉; 目标跟踪; 相关滤波; 自适应特征融合; 上下文感知

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.1610011

Object Tracking Based on Adaptive Feature Fusion and Context-Aware

Ji Yuanfa^{1,2}, He Chuanji^{1,2}, Sun Xiyan^{1,2*}, Guo Ning^{1,2}

¹Guangxi Key Laboratory of Precision Navigation Technology and Application, Guilin University of Electronic Technology, Guilin, Guangxi 541004, China;

²National & Local Joint Engineering Research Center of Satellite Navigation and Location Service, Guilin, Guangxi 541004, China

Abstract The description of a target's appearance greatly influences the performance of a correlation filter tracker. It is difficult to obtain an accurate description of target's appearance using a single feature. Therefore, the target appearance description based on multiple features can improve the tracking performance in complex scenes. For robust object tracking in complex scenes, we propose an object tracking algorithm based on multiple features with adaptive fusion and context-aware. First, we introduced a context-aware framework and extracted single-layer convolution features from four context image patches around the target to establish the background information. As a single feature may not accurately describe the target appearance, two correlation filters were used for feature extraction. The first filter extracted three-layer convolution features as deep features through a convolutional neural network, and the second filter extracted information from the directional gradient histogram and color histogram to obtain shallow features. Then, the deep and shallow features were adaptively fused. Finally, the average peak-to-correlation energy was used to evaluate the confidence of the response and we decided whether to update the model. The proposed algorithm was evaluated on the OTB-2013 benchmark, and the results show that it achieves excellent performance regarding accuracy and success rate and shows superior tracking performance compared with other state-of-the-art tracking algorithms.

收稿日期: 2020-07-31; 修回日期: 2020-11-10; 录用日期: 2020-12-22

基金项目: 国家自然科学基金(61561016, 61861008)、桂林电子科技大学研究生教育创新计划(2020YCXS030)

通信作者: *sunxiyan1@163.com

Key words image processing; machine vision; object tracking; correlation filtering; adaptive feature fusion; context-aware

OCIS codes 100.3008; 150.0155; 100.4999

1 引言

视觉跟踪一直是计算机视觉领域关注的重点问题。近年来,基于相关滤波的视觉跟踪算法发展迅速,在跟踪速度和精度上都展现了一定的优势。然而,目标跟踪的研究仍然存在一定的困难,目标遮挡、快速运动、光照变化等外界干扰因素都直接影响着跟踪算法的性能。数据集的出现给评估跟踪算法的性能带来了很大的帮助,目前已有数据集^[1-2]可供测试。

基于相关滤波和基于深度学习的方法是当前主流的目标跟踪算法,其中相关滤波凭借在频域的快速计算特性带来的速度优势,更是成为了当前目标跟踪领域的研究热点之一^[3-6]。Bolme 等^[5]首先将相关滤波算法引入到目标跟踪领域中,提出了一种最小输出平方误差滤波器(MOSSE)算法,但灰度特征并不能准确地描述目标的外观。Henriques 等^[7]通过改进 MOSSE 算法的核函数,提出了核函数循环结构(CSK)算法。该算法通过循环移位方式完成样本的密集采样,解决了训练样本不足的问题,但采用单通道的灰度特征,特征描述能力有限。随后 Henriques 等^[8]又提出了核相关滤波(KCF)算法,该算法将核函数引入岭回归中,将单通道的灰度特征扩展为多通道的梯度直方图(HOG)特征。Possegger 等^[9]使用颜色直方图特征来描述目标外观,取得了良好的效果。Danelljan 等^[10]利用颜色属性(CN)特征扩展 CSK 算法,通过主成分分析(PCA)降维来降低运算量,提高了跟踪精度。上述方法只利用单一特征对目标进行外观描述,在复杂场景下目标的辨识度差并且容易受到干扰,因此,采用多特征描述目标外观的方法被引入到相关滤波算法中。Li 等^[11]提出了一种有效的尺度估计策略,融合 CN 特征与 HOG 特征来提升算法性能。Bertinetto 等^[12]提出了 Staple 算法,将 HOG 特征和颜色直方图特征线性相加,提升了算法性能。沈秋等^[13]将 HOG 特征、CN 特征和局部二值模式(LBP)特征相结合,其中最大响应所在的位置即为跟踪目标。

深度特征包含高层语义,鲁棒性很强,对旋转和变形等外观变化不敏感,能够更好地描述目标的外

观,大量基于深度特征的跟踪算法被相继提出^[14-20]。Wang 等^[16]发现不同的卷积层特征对跟踪问题的影响不同,高层的深度特征具有更好的鲁棒性,低层的深度特征则提供了更多的局部信息,结合两种层次深度特征的跟踪算法能够有效提升跟踪性能。Ma 等^[18]利用分层卷积神经网络特征描述目标外观,通过线性加权融合各分层响应,实现了目标的准确定位。Danelljan 等^[21]结合深度特征、CN 特征、HOG 特征,提出了连续卷积相关滤波跟踪(CCOT)算法,显著提升了跟踪精度。文献[22]通过对卷积操作进行因式分解来减少模型参数,简化训练集生成的同时保证训练样本的多样性,采用模型稀疏更新策略来提升跟踪器性能。为了充分发挥深度特征的强大性能,Bhat 等^[23]将深度特征和浅层特征分开训练,深度特征负责鲁棒性,浅层特征负责准确性,并提出了一种全新的深、浅层特征融合的方式,有效提升了跟踪算法的性能。

传统的相关滤波跟踪算法存在一定的局限性,比如尺度变化和边界效应。Danelljan 等^[24-25]考虑到特征值的多重维度,提出了平移滤波结合尺度滤波方式的判别式尺度空间跟踪器(DSST)。在解决边界效应方面,Danelljan 等^[26]使用 Gauss-Seidel 函数来求解滤波器,加入了空间正则来解决边界效应;Mueller 等^[27]提出了上下文感知(context-aware)框架,该框架能够显式地学习目标周围的背景信息,有效地减轻了边界效应,同时可以广泛地应用到基于相关滤波的跟踪算法上。

本文通过自适应地结合深、浅层互补特征来提高跟踪性能。考虑到边界效应的影响,所提算法引入上下文感知框架,提取目标周围 4 个图像块的 conv4-4 层卷积特征作为背景信息。与仅使用单一特征的算法相比,所提算法包含两个相关滤波器,一个滤波器通过卷积神经网络提取三层卷积特征作为深度特征,另一个滤波器提取 HOG 和颜色直方图信息作为浅层特征。其次,在特征响应融合阶段,采用了一种新的多特征融合策略来有效地结合深、浅层互补特征。最后,使用平均峰值相关能量(APCE)指数来评估响应图的置信度和跟踪结果的可靠性,并决定是否更新模型。所提算法在 OTB-2013 数据集上进行测试,结果表明,所提算法在准

确率和成功率方面都有出色的表现。

2 相关滤波跟踪中的特征应用

特征在视觉跟踪中起着非常重要的作用,使用合适的特征来描述目标外观可以有效地提高跟踪性能。MOSSE^[5]和CSK^[7]使用灰度特征来描述目标的外观,但是特征描述能力有限。KCF^[8]将单通道的灰度特征扩展为多通道的HOG特征,提高了对目标外观的描述能力。DAT^[9]使用颜色直方图特征来描述目标的外观,取得了很好的效果。文献[10]利用CN特征扩展了CSK算法,通过主成分分析降维,提高了跟踪精度。深度特征包含高级语义信息,鲁棒性强,能较好地描述目标的外观。文献[16]发现不同卷积层的深度特征对跟踪问题的影响不同,高层的深度特征具有更好的鲁棒性,而低层的深度特征提供更多的局部信息,结合多层深度特征的

跟踪算法具有更好的跟踪性能。Staple算法^[12]对HOG特征和颜色直方图特征进行线性结合,提高了算法的性能。UPDT算法^[23]基于ECO^[22],自适应地结合了卷积特征和浅层特征的响应,取得了优秀的跟踪效果。

3 所提算法

3.1 算法概述

提出一种有效的视觉跟踪算法。首先引入了上下文感知框架,提取目标周围4个图像块的conv4-4层卷积特征作为背景信息。与只使用深度特征的算法相比,所提算法还提取颜色直方图与HOG特征来和深度特征实现特性互补。然后,采用一种新的自适应特征融合方法来有效结合深、浅层特征。最后,利用APCE来评估跟踪结果的可靠性,并决定是否对模型进行更新。所提算法的整体流程如图1所示。

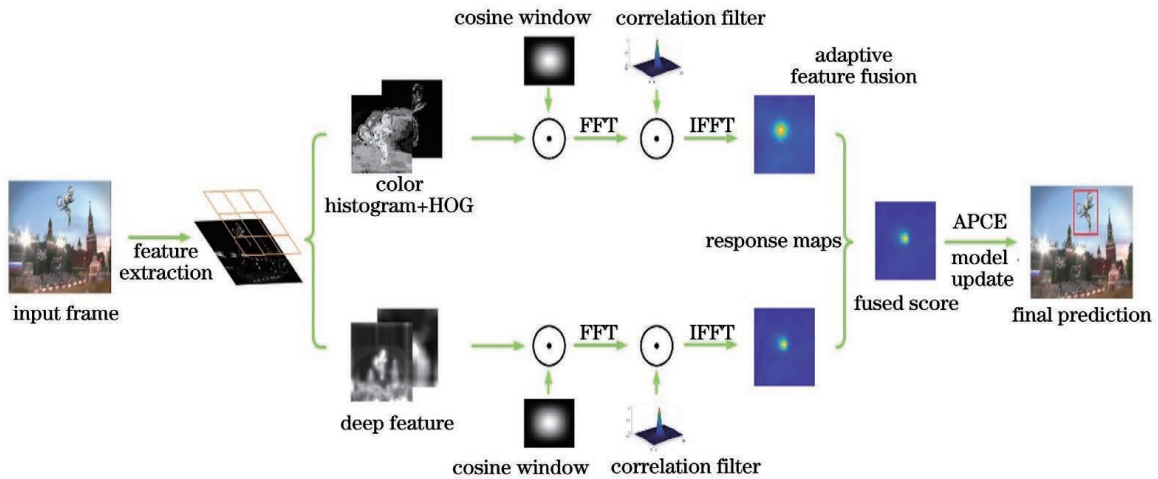


图1 所提算法的流程

Fig. 1 Flowchart of the proposed algorithm

3.2 上下文感知框架

目标周围的背景信息对跟踪算法的性能有非常重要的影响。由于循环样本的特性,相关滤波跟踪器容易产生边界效应,余弦窗可以有效地限制边界效应但却减少了背景信息,当目标发生形变时可能导致跟踪失败。针对这个问题,引入了上下文感知框架。在实际应用中,考虑到对每个上下文图像块提取多层卷积特征会大幅降低跟踪速度,而不会显著提高跟踪性能,因此只提取目标周围4个上下文图像块的conv4-4层卷积特征作为负样本,并将其引入到模型学习过程中。在每一帧图像中,所提算法在跟踪对象 $\mathbf{n}_0 \in \mathbb{R}^n$ 周围提取 k 个上下文背景图像块 $\mathbf{n}_i \in \mathbb{R}^n$,相应的循环矩阵为 $\mathbf{N}_0 \in \mathbb{R}^{n \times n}$ 和

$\mathbf{N}_i \in \mathbb{R}^{n \times n}$ 。将采集到的上下文背景图像块作为正则化项引入到岭回归方程中,得到

$$\min_{\mathbf{w}} \|\mathbf{N}_0 \mathbf{w} - \mathbf{y}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2^2 + \lambda_2 \sum_{i=1}^k \|\mathbf{N}_i \mathbf{w}\|_2^2, \quad (1)$$

式中: \mathbf{w} 为通过学习得到的相关滤波器;矩阵 \mathbf{N}_0 包含了矢量图像 \mathbf{n}_0 的所有循环移位;回归目标 \mathbf{y} 为二维高斯向量化后的图像; λ_2 为防止背景样本回归值趋于零的正则化参数; λ_1 为控制过拟合的正则化参数。对原始目标区域与加入上下文信息的背景样本进行叠加,优化(1)式得到

$$f(\mathbf{w}, \mathbf{B}) = \|\mathbf{B}\mathbf{w} - \bar{\mathbf{y}}\|_2^2 + \lambda_1 \|\mathbf{w}\|_2^2, \quad (2)$$

式中: \mathbf{B} 为包括目标及上下文区域的特征矩阵; $\bar{\mathbf{y}}$ 为

对应目标样本与背景样本的响应矩阵。目标函数 $f(\mathbf{w}, \mathbf{B})$ 是凸函数, 可以通过设置梯度为零来最小化得到

$$\mathbf{w} = (\mathbf{B}^T \mathbf{B} + \lambda_1 \mathbf{I})^{-1} \mathbf{B}^T \bar{\mathbf{y}}, \quad (3)$$

式中: \mathbf{I} 为单位矩阵。将循环矩阵转换到频域上进行对角化, 可得在傅里叶域的封闭解:

$$\hat{\mathbf{w}} = \frac{\hat{\mathbf{n}}_0^* \odot \hat{\mathbf{y}}}{\hat{\mathbf{n}}_0^* \odot \hat{\mathbf{n}}_0 + \lambda_1 \mathbf{I} + \lambda_2 \sum_{i=1}^k \hat{\mathbf{n}}_i^* \odot \hat{\mathbf{n}}_i}, \quad (4)$$

式中: $\hat{\mathbf{n}}_0^*$ 为矢量图像 \mathbf{n}_0 傅里叶变换的共轭形式; $\hat{\mathbf{y}}$ 为回归目标 \mathbf{y} 的傅里叶变换。(3) 式中原始域中的解与标准岭回归问题的解形式完全相同, 对偶域中的解为

$$\boldsymbol{\alpha} = (\mathbf{B}\mathbf{B}^T + \lambda_1 \mathbf{I})^{-1} \bar{\mathbf{y}}, \boldsymbol{\alpha} \in \mathbb{R}^{(k+1)n}. \quad (5)$$

利用循环矩阵的特殊性质得到

$$\hat{\boldsymbol{\alpha}} = \begin{bmatrix} \text{diag}(\mathbf{d}_{00}) & \cdots & \text{diag}(\mathbf{d}_{0k}) \\ \vdots & & \vdots \\ \text{diag}(\mathbf{d}_{k0}) & \cdots & \text{diag}(\mathbf{d}_{kk}) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{y}} \\ \vdots \\ 0 \end{bmatrix}, \quad (6)$$

式中: \mathbf{d}_{jl} 为给定的向量, $j, l \in \{1, \dots, k\}$ 。

$$\begin{cases} \mathbf{d}_{00} = \hat{\mathbf{n}}_0 \odot \hat{\mathbf{n}}_0^* + \lambda_1 \mathbf{I} \\ \mathbf{d}_{jj} = \lambda_2 (\hat{\mathbf{n}}_j \odot \hat{\mathbf{n}}_j^*) + \lambda_1 \mathbf{I}, j \neq 0. \\ \mathbf{d}_{jl} = \sqrt{\lambda_2} (\hat{\mathbf{n}}_j \odot \hat{\mathbf{n}}_l^*), j \neq l \end{cases} \quad (7)$$

(7) 式图像之间的相关性可以使用线性核进行加速运算^[8], 引入上下文背景图像块, 检测公式最终为

$$\hat{\mathbf{r}}_d = \hat{\mathbf{z}} \odot \hat{\mathbf{n}}_0^* \odot \hat{\boldsymbol{\alpha}}_0 + \sqrt{\lambda_2} \sum_{i=1}^k \hat{\mathbf{z}} \odot \hat{\mathbf{n}}_i^* \odot \hat{\boldsymbol{\alpha}}_i, \quad (8)$$

式中: \mathbf{r} 为搜索窗口中的最大响应位置; \mathbf{z} 为上下文图像块里的搜索窗口; $\boldsymbol{\alpha}$ 为对偶变量。

3.3 多特征提取

目标外观的描述对相关滤波跟踪器的性能有着很大的影响, 单一特征难以准确描述目标外观, 基于多特征的目标外观描述在复杂场景中有着更加优异的性能。深度特征包含高层次的语义信息, 对旋转、变形等外观变化不敏感, 具有较强的鲁棒性。采用 VGG-NET-19 进行深度特征提取。选择包含更多高层语义信息的 conv5-4, 包含更多细节信息的 conv3-4 和 conv4-4 层卷积特征, 经过加权融合后作为最终使用的深度特征。

浅层特征主要是手工制作的特征, 例如 RGB 像素、HOG 和颜色特征, 包含纹理、颜色等细节信息, 空间分辨率高, 适合高精度定位。从原始图像中提取颜色直方图和 HOG 特征作为浅层特征, 并与经

过卷积神经网络提取的深度特征相结合来描述目标的外观。

3.4 自适应特征融合

深度特征编码高层语义信息, 对外观变化不敏感, 可以用于粗定位, 而浅层特征细节分辨率较高, 适合精确定位。UPDT^[23] 将两种特征分开处理, 深度特征负责鲁棒性, 浅层特征强调准确性, 两种特征自适应融合, 实现特征互补。所提算法提取三层卷积特征作为深度特征, 颜色直方图特征和 HOG 特征作为浅层特征, 两种特征分别训练相关滤波器, 建立两个独立的外观模型, 采用自适应特征融合策略结合两种特征的响应图, 融合后的特征是两种特征的加权组合。

$$y_\beta(t) = \beta_d y_d(t) + \beta_s y_s(t), \quad (9)$$

式中: y_d 为深层特征分数; y_s 为浅层特征分数; y_β 为两种分数加权得到的总分, 深、浅分数的权重 $\beta = (\beta_d, \beta_s)$ 。

UPDT 指出响应图能够反映预测目标的准确性和鲁棒性。准确性与预测目标周围响应峰值的锐利程度有关, 主峰越尖锐则精度越强。鲁棒性与主峰到干扰峰的距离有关, 主峰到次峰的距离越大则鲁棒性越强。为了评价预测目标的可靠性, 采用一种预测质量评价方法。

$$\xi_{t^*} \{y\} = \min_t \frac{y(t^*) - y(t)}{\Delta(t - t^*)}, \quad (10)$$

式中: y 为图像坐标上的得分函数; $y(t) \in \mathbb{R}$ 是位置 $t \in \mathbb{R}^2$ 的目标预测分数; t^* 为候选的预测目标。距离测量函数 Δ 定义为

$$\Delta(\tau) = 1 - e^{-\frac{\kappa}{2} |\tau|^2}, \quad (11)$$

式中: κ 是控制 $\Delta(\tau) \rightarrow 1$ 变化速率的参数。基于(9)式, 共同估计分数权重 β 和目标状态 t^* , 最大化质量评估(10)式, 通过最小化损失函数得到

$$\begin{cases} \text{minimize: } L_{t^*}(\beta) = -\xi_{t^*} \{y_\beta\} + \mu(\beta_d^2 + \beta_s^2) \\ \text{subject to: } \beta_d + \beta_s = 1, \beta_d \geq 0, \beta_s \geq 0 \end{cases}. \quad (12)$$

为了优化(12)式, 引入松弛变量 $\xi = \xi_{t^*} \{y_\beta\}$, 等价于求解最小化问题。

$$\begin{cases} \text{minimize: } L_{t^*}(\xi, \beta) = -\xi + \mu(\beta_d^2 + \beta_s^2) \\ \text{subject to: } \beta_d + \beta_s = 1, \beta_d \geq 0, \beta_s \geq 0 \\ y_\beta(t^*) - \xi \Delta(t^* - t) \geq y_\beta(t), \forall t \in \Omega \end{cases}, \quad (13)$$

式中: μ 为控制正则项的参数, 用来对权重的较大偏差进行惩罚。在实际操作中, 所提算法分别从深层

分数和浅层分数中寻找局部极大值,对局部极大值按照大小进行排序筛选后作为有限候选状态 Ω ,通过(13)式优化每个状态 $t^* \in \Omega$,然后选择具有最低总体损失的候选状态 t^* 作为最终的预测结果,并得到相应的权重系数 $\beta = (\beta_d, \beta_s)$ 。

3.5 模型更新

在跟踪过程中,算法不可避免地会受到各种因素的干扰,比如目标丢失、背景遮挡或模糊等,使用错误的信息更新模型可能会导致跟踪漂移甚至失败。为了解决问题,采用 APCE 来评估响应的置信度。该指标能反映跟踪结果的可信度和响应图的波动程度。APCE 的表达式为

$$E_{APC} = \frac{|\mathbf{F}_{\max} - \mathbf{F}_{\min}|^2}{\text{mean} \left[\sum_{w,h} (\mathbf{F}_{w,h} - \mathbf{F}_{\min})^2 \right]}, \quad (14)$$

式中: \mathbf{F}_{\max} 为最大响应值; \mathbf{F}_{\min} 为最小响应值; $\mathbf{F}_{w,h}$ 为位置 (w, h) 上的响应值。APCE 值和响应如图 2 所示。当目标发生形变或存在背景干扰时,响应图产生剧烈波动并出现多峰干扰,同时 APCE 值处于较低的状态。当目标没有受到干扰时,响应图会有一个锐利且明确的峰值,同时 APCE 处于较高的状态。因此,当 APCE 值明显减小时不更新模型,跟踪器只在 APCE 值大于阈值的时候才更新模型,这样可以有效避免模型漂移的问题。

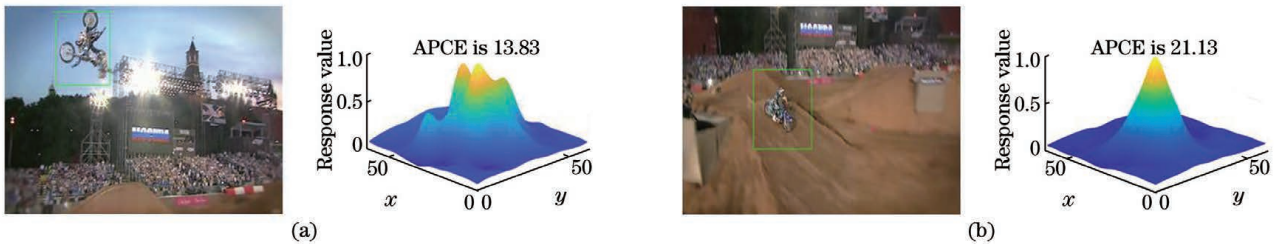


图 2 不同场景下的 APCE 和响应值。(a)场景 1;(b)场景 2

Fig. 2 APCE value and response value at different scenes. (a) Scene 1; (b) scene 2

4 实验与分析

4.1 实验环境及配置

实验硬件环境为 Win10 操作系统, Intel Core i7-8750H(2.20 GHz)处理器, 8 GB 内存, MATLAB R2018a。使用 OTB-2013 中的 49 个视频序列来测试所提算法, 实验测试序列中包含了目标跟踪中常见的复杂场景, 包括快速运动、目标遮挡、背景杂波、运动模糊和尺度变化等。深度特征、HOG 特征和颜色直方图特征的学习率分别为 0.01、0.04 和 0.01, λ_1 为 0.00001, λ_2 为 0.2, APCE 的阈值为 0.7。

根据 OTB 数据集的评估方法, 采用一遍评估 (OPE) 法, 从跟踪精度和跟踪成功率两个方面对所有跟踪器进行评估。跟踪精度指预测目标中心位置误差小于某一阈值的帧数占所有帧数的百分比; 跟踪成功率表示预测框的重叠率, 即跟踪重叠率大于阈值的帧数占所有帧数的百分比; 重叠率指预测边界框与实际边界框的重叠区域与它们并集的总面积之比。对所提算法与 HCF^[18]、SRDCF^[26]、Staple^[12]、LCT^[28]、SiamFC^[29]、RPT^[30]、SAMF^[11]、KCF^[8]、DSST^[24] 算法进行比较。

4.2 特征融合分析

深度特征包含高层语义信息, 对旋转和变形等

外观变化不敏感, 具有很强的鲁棒性, 而颜色直方图和 HOG 特征空间分辨率高、准确性很强, 两种深、浅层特征相结合能够更加有效地描述目标外观。提取颜色直方图特征和 HOG 特征作为浅层特征, 以 OTB-2013 数据集中 MotorRolling 序列中的第 48 帧图像为例, 提取得到的颜色直方图如图 3 所示。

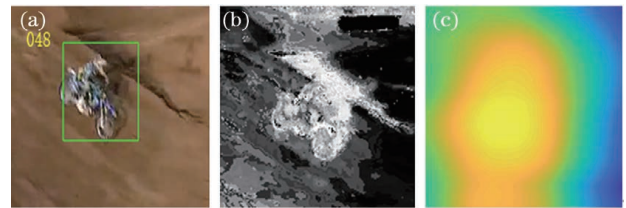


图 3 颜色直方图和响应图。(a)原图;(b)颜色直方图;(c)颜色特征的响应图

Fig. 3 Color histogram and response map. (a) Original image; (b) color histogram; (c) response map of color feature

图 3(b) 是特征提取得到的颜色直方图, 保留了较高的分辨率。以 OTB-2013 数据集中的 Soccer 序列为例, 对所提自适应特征融合算法进行实验分析。所提算法通过找到深层和浅层得分中具有最低总体损失的候选状态来计算权重值。图 4 显示了 Soccer 序列中每一帧图像的深、浅得分的自适应权重值。特征权重值在跟踪过程中不断变化。总体来

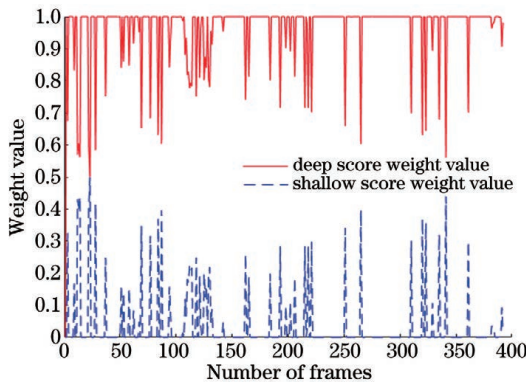


图 4 深层和浅层分数的自适应权重值

Fig. 4 Adaptive weight values for deep and shallow scores 看,深度特征的权重值远大于浅层特征的权重值,这是因为深度特征的强大性能在跟踪过程中起着决定性的作用。深度特征包含高层语义信息,对目标外

观变化不敏感,具有较强的鲁棒性,而 Soccer 视频序列中存在复杂的背景和颜色干扰;浅层特征响应图在跟踪过程中存在着巨大的波动,在这种情况下浅层分数是不可靠的。因此深层分数的权重值会高于浅层分数的权重值。

为了验证所提自适应特征融合策略的跟踪性能,在 OTB-2013 测试集中进行 3 种算法的性能对比。3 种算法包括所提使用深度特征、颜色直方图和 HOG 特征并进行自适应特征融合的 OUR 算法,分别采用深度特征 (DEEP)、颜色直方图和 HOG 特征 (CN+HOG) 的 OUR_DEEP 和 OUR_CN+HOG 算法。OUR_DEEP 和 OUR_CN+HOG 算法的实验条件和参数均与所提算法一致。3 种算法的对比实验结果如图 5 所示。

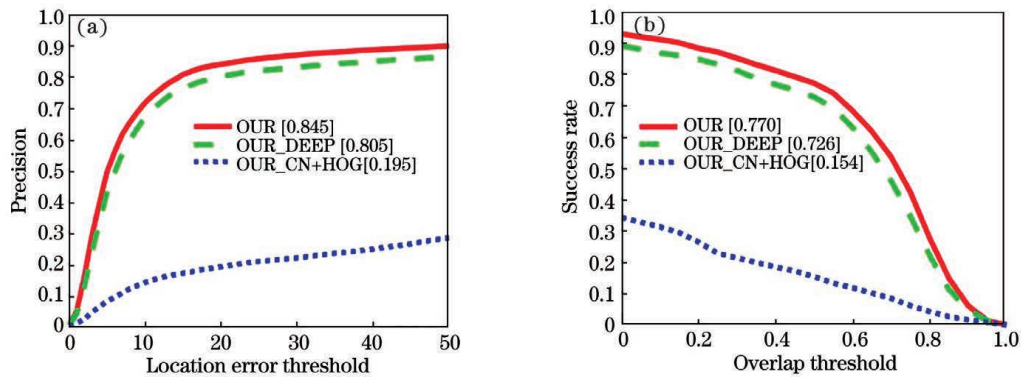


图 5 3 种算法的跟踪性能对比。(a)精度;(b)成功率

Fig. 5 Comparison of tracking performance of three algorithms. (a)Precision; (b)success rate

从图 5 可以看出,与只采用单一深度特征、颜色直方图和 HOG 特征的 OUR_DEEP 和 OUR_CN+HOG 算法相比,所提采用自适应特征融合策略的算法在跟踪精度上分别提高了 4.9% 和 333%,在跟踪成功率上分别提高了 6% 和 400%,这说明自

适应特征融合策略可以有效提升算法的跟踪性能。

4.3 定量分析

所提算法在 OTB-2013 数据集上的实验结果如图 6~8 所示。图 6 显示所提算法 (OUR) 和其他优秀的跟踪算法在 OTB-2013 数据集上的精度图和成

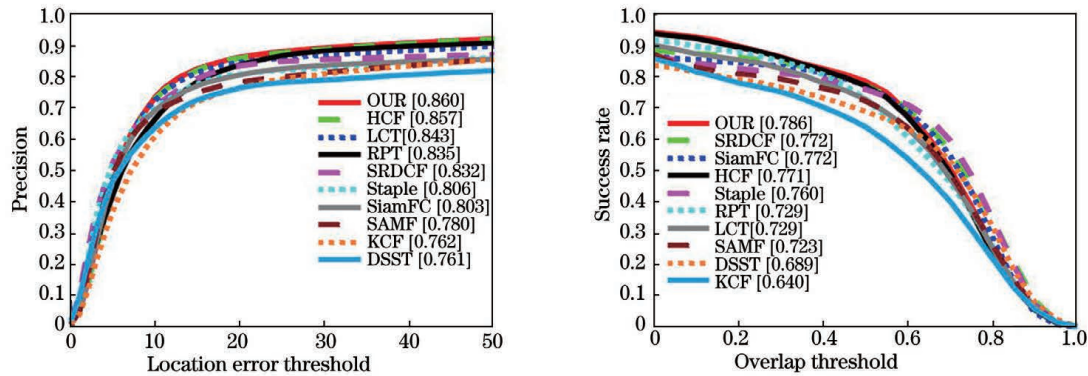


图 6 在 OTB-2013 数据集上 10 种跟踪算法的精度和成功率比较。(a)精度;(b)成功率

Fig. 6 Comparison of the precision and success rate of 10 tracking algorithms on OTB-2013 benchmark. (a) Precision; (b) success rate

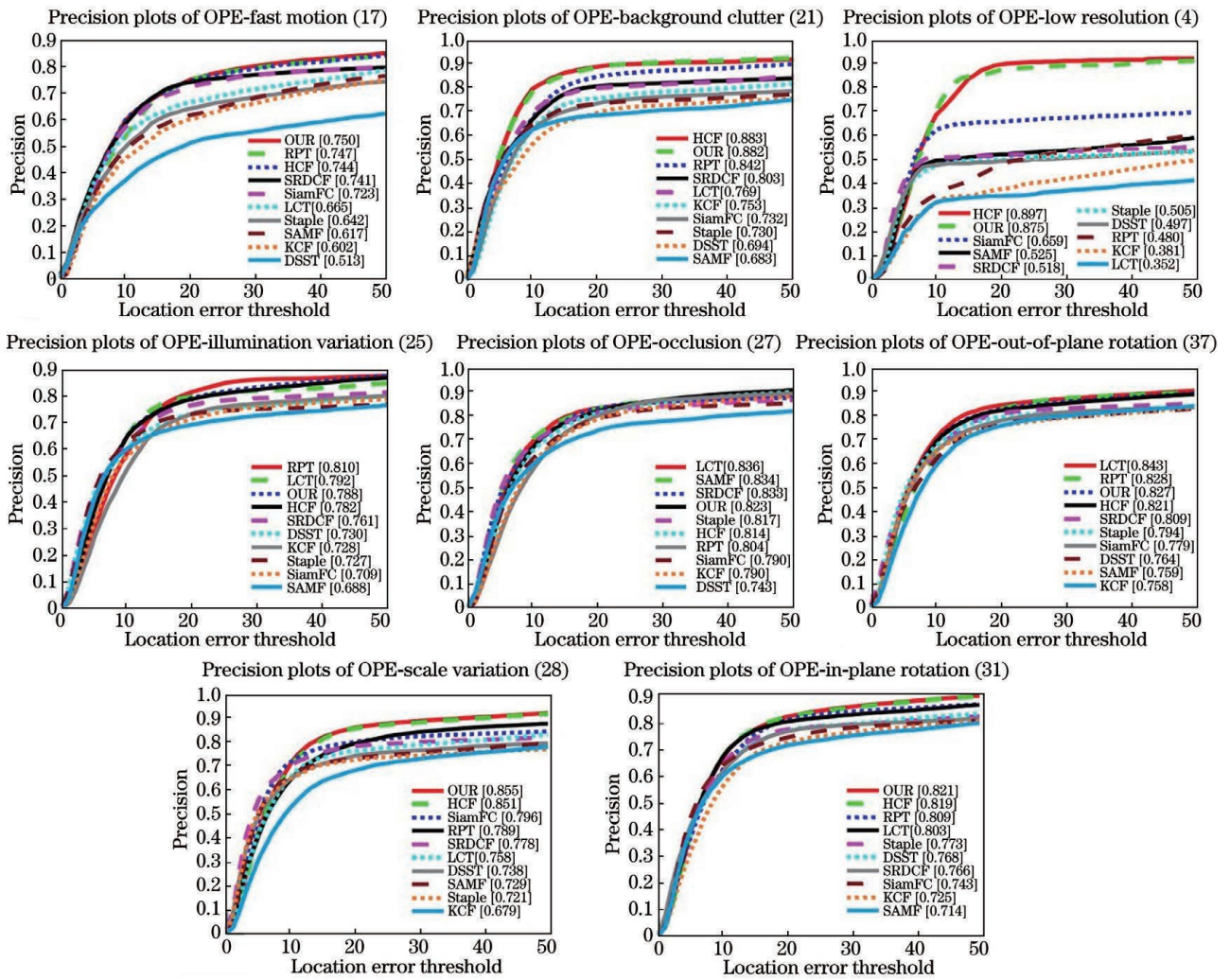


图 7 在 OTB-2013 数据集上 8 种属性序列的跟踪精度比较

Fig. 7 Comparison of tracking precision of 8 attribute sequences on OTB-2013 benchmark

功率图。可以看到:所提算法在精度和成功率图上均处于第一位;所提算法的跟踪精度能够达到 0.860,相比排名第二的 HCF 算法精度提高了 0.3%,所提算法的成功率可以达到 0.786,比 HCF 提高了 1.9%;所提算法的跟踪精度相比 SRDCF、LCT、SiamFC 和 Staple 算法分别提高了 3.3%、2%、7%和 6.6%;在跟

踪成功率方面,所提算法与 SRDCF、LCT、SiamFC 和 Staple 相比,成功率分别提高了 1.8%、7.8%、1.8%和 3.4%。在跟踪速度方面,如表 1 所示,所提算法只有 5 frame/s,这是因为引入了上下文感知框架,提取了目标周围上下文图像块的 conv4-4 层卷积特征,提高了跟踪性能,但大幅降低了速度。

表 1 10 种跟踪算法的跟踪速度

Table 1 Tracking speed of 10 tracking algorithms

Parameter	OUR	HCF	SRDCF	Staple	LCT	SiamFC	RPT	SAMF	KCF	DSST
Tracking speed / (frame · s ⁻¹)	5	10	4	75	28	55	6	8	172	21

图 7 和图 8 分别显示了所提算法和其他优秀的跟踪算法在 OTB-2013 数据集上的 8 个属性的精度和成功率,包括快速运动、背景杂波、低分辨率、遮挡、平面外旋转、平面内旋转、比例变化和光照变化。

所提算法在大多数属性上都处于第一位。具体来说,所提算法在快速运动、尺度变化和平面内旋转属性的跟踪精度分别达到了 0.750、0.855 和 0.821,高于其他算法。在跟踪成功率方面,所提算法在低分辨率下的跟踪精度可以达到 0.825,相比

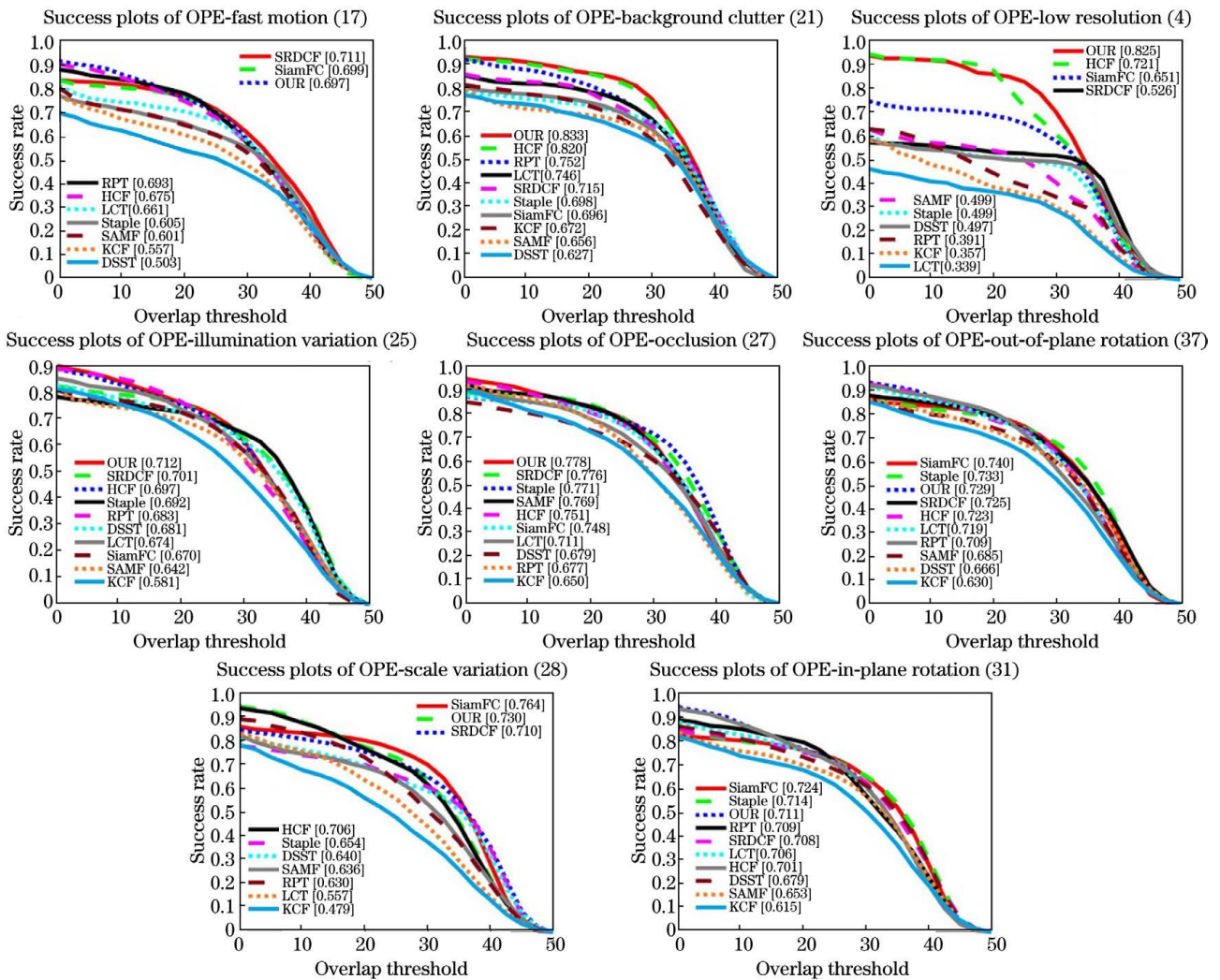


图 8 在 OTB-2013 数据集上 8 种属性序列的跟踪成功率比较

Fig. 8 Comparison of tracking success rates of 8 attribute sequences on OTB-2013 benchmark

HCF、SRDCF 和 SiamFC 算法分别提高了 14.4%、56.8%和 26.7%。在光照变化方面,所提算法的成功率可以达到 0.712,与 HCF、SRDCF 和 Staple 相比,成功率分别提高了 2.1%、1.5%和 2.8%。结果表明,所提算法在大多数具有挑战性的属性中都有着优秀的性能。

4.4 定性分析

使用 OTB-2013 数据集中的 Ironman、Shaking、Soccer 和 Skiing 序列对所提算法进行定性分析,这些视频序列包含了尺度变化、光照变化、平面外旋转、变形和背景杂波等属性。图 9 显示了 10 种算法的跟踪结果。

在 Ironman 序列中,目标的姿态在运动过程中不断变化,背景和光照在运动过程中剧烈变化,只有所提算法和 HCF 算法可以连续准确地跟踪目标,其他算法都存在严重的偏差。在 Shaking 序列中,

第 76 帧图像存在剧烈的光照变化,Staple 和 KCF 算法存在严重的漂移现象,而所提算法仍能保持对目标的准确跟踪。Soccer 序列的跟踪目标不断地被颜色和背景杂波所干扰,但所提算法仍然可以准确地跟踪目标,这是因为所提算法结合了多种特征来有效地描述目标外观,可以实现对目标的精确跟踪。

在 Ironman 和 Shaking 的场景中,背景是复杂多变的。在 Ironman 序列的第 50 帧中有几个相似的目标,其他算法都已跟踪失败,只有所提算法、HCF 和 SiamFC 能够准确地跟踪目标。在 Shaking 序列的第 85 帧,背景存在光照变化,目标周围存在相似的人脸干扰。Staple 算法产生了严重的偏差,只有所提算法、HCF、SiamFC、RPT、LCT 和 DSST 能够准确跟踪目标。这是因为所提算法使用平均峰值相关能量来判断跟踪结果的可靠性。当存在多个



图 9 10 种跟踪算法在 OTB-2013 数据集上的定性比较

Fig. 9 Qualitative comparison of 10 tracking algorithms on OTB-2013 benchmark

可疑目标时,响应图会产生剧烈波动,模型不会更新,提高了算法的抗干扰能力。在 Skiing 序列中,目标快速移动,同时伴随着姿态旋转变形和背景剧烈变化。其他算法从第 11 帧开始出现严重偏差,跟踪失败,只有所提算法和 HCF 算法能够持续准确地跟踪目标,这说明所提算法具有很强的稳定性。

5 结 论

传统的相关滤波跟踪算法使用的手工特征并不能很好地描述跟踪目标的外观,存在较为严重的边界效应的同时也缺乏合适的特征融合策略。所提算法通过引入上下文感知框架来减少边界效应,提取多层卷积特征作为深度特征,HOG 和颜色直方图信息作为浅层特征,采用自适应特征融合策略,并利用平均峰值相关能量来评估跟踪结果的可靠性,再决定是否更新模型。所提算法在 OTB-2013 数据集上进行了测试,并与近年来其他优秀的相关滤波跟踪算法进行了比较。实验结果表明,所提算法具有优秀的跟踪性能。

参 考 文 献

- [1] Smeulders A W M, 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.
- [2] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 2411-2418.
- [3] Galoogahi H K, Sim T, Lucey S. Multi-channel correlation filters [C] // 2013 IEEE International Conference on Computer Vision, December 1-8, 2013, Sydney, NSW, Australia. New York: IEEE Press, 2013: 3072-3079.
- [4] Huang C, Lucey S, Ramanan D. Learning policies for adaptive tracking with deep feature cascades[C]// 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 105-114.
- [5] 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, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE Press, 2010: 2544-2550.
- [6] Shen Y L, Wu Z D, Zhao R J, et al. Long-term object tracking based on model updating and fast re-detection[J]. *Acta Optica Sinica*, 2020, 40(3): 0315002.
沈玉玲, 伍忠东, 赵汝进, 等. 基于模型更新与快速重检测的长时目标跟踪[J]. *光学学报*, 2020, 40(3): 0315002.
- [7] Henriques J F, Caseiro R, Martins P, et al. Exploiting the circulant structure of tracking-by-detection with kernels[M]//Fitzgibbon A, Lazebnik S, Perona P, et al. *Computer vision-ECCV 2012*. Lecture notes in computer science. Heidelberg: Springer, 2012, 7575: 702-715.
- [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] Possegger H, Mauthner T, Bischof H. In defense of color-based model-free tracking [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 2113-2120.
- [10] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 1090-1097.
- [11] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration[M]//Agapito L, Bronstein M M, Rother C. *Computer vision-ECCV 2014 workshops*. Lecture notes in computer science. Cham: Springer, 2015, 8926: 254-265.
- [12] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: complementary learners for real-time tracking[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1401-1409.
- [13] Shen Q, Yan X L, Liu L F, et al. Multi-scale correlation filtering tracker based on adaptive feature selection[J]. *Acta Optica Sinica*, 2017, 37(5): 0515001.
沈秋, 严小乐, 刘霖枫, 等. 基于自适应特征选择的多尺度相关滤波跟踪[J]. *光学学报*, 2017, 37(5): 0515001.
- [14] Qi Y K, Zhang S P, Qin L, et al. Hedged deep tracking[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 4303-4311.
- [15] Wang L J, Ouyang W L, Wang X G, et al. STCT: sequentially training convolutional networks for visual tracking[C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1373-1381.
- [16] Wang L J, Ouyang W L, Wang X G, et al. Visual tracking with fully convolutional networks[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 3119-3127.
- [17] Hong S, You T, Kwak S, et al. Online tracking by learning discriminative saliency map with convolutional neural network[J]. *Journal of Machine Learning Research*, 2015: 597-606.
- [18] Ma C, Huang J B, Yang X K, et al. Hierarchical convolutional features for visual tracking[C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 3074-3082.
- [19] Chen F L, Ding Q H, Chang Z, et al. Multi-scale kernel correlation filter algorithm for visual tracking based on the fusion of adaptive features[J]. *Acta Optica Sinica*, 2020, 40(3): 0315001.
陈法领, 丁庆海, 常铮, 等. 自适应特征融合的多尺度核相关滤波目标跟踪[J]. *光学学报*, 2020, 40(3): 0315001.
- [20] Wang H Y, Wang L, Yin W R, et al. Multi-scale correlation filtering visual tracking algorithm combined with target detection[J]. *Acta Optica Sinica*, 2019, 39(1): 0115004.
王红雨, 汪梁, 尹午荣, 等. 结合目标检测的多尺度相关滤波视觉跟踪算法[J]. *光学学报*, 2019, 39(1): 0115004.
- [21] Danelljan M, Robinson A, Khan F S, et al. Beyond correlation filters: learning continuous convolution operators for visual tracking[M]//Leibe B, Matas J, Sebe N, et al. *Computer vision-ECCV 2016*. Lecture notes in computer science. Cham: Springer, 2016, 9909: 472-488.
- [22] Danelljan M, Bhat G, Khan F S, et al. ECO: efficient convolution operators for tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6931-6939.
- [23] Bhat G, Johnander J, Danelljan M, et al. Unveiling the power of deep tracking[M]//Ferrari V, Hebert

- M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11206: 493-509.
- [24] Danelljan M, Häger G, Shahbaz Khan F, et al. Accurate scale estimation for robust visual tracking [J]. Proceedings of the British Machine Vision Conference 2014, 2014: 65.1-65.11.
- [25] 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.
- [26] Danelljan M, Häger G, Khan F S, et al. Learning spatially regularized correlation filters for visual tracking[C]//2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 4310-4318.
- [27] Mueller M, Smith N, Ghanem B. Context-aware correlation filter tracking[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 1387-1395.
- [28] Ma C, Yang X K, Zhang C Y, et al. Long-term correlation tracking [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 5388-5396.
- [29] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.
- [30] Li Y, Zhu J K, Hoi S C H. Reliable patch trackers: robust visual tracking by exploiting reliable patches [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 353-361.