

基于高置信度更新的多滤波器协同跟踪算法

张超溢, 彭力, 贾天豪, 闻继伟*

江南大学物联网工程学院轻工过程先进控制教育部重点实验室, 江苏 无锡 214122

摘要 提出一种基于高置信度模型更新策略的多相关滤波器协同跟踪算法。利用卷积网络结构 VGG-Net-19 提取目标周围区域的多层卷积特征, 构造深度滤波器, 以自适应的特征融合策略实现目标初定位; 建立尺度滤波器以检测目标的尺寸变化; 利用主次峰坡度比作为跟踪置信度指标, 设计一种高置信度下的模型更新策略; 当跟踪置信度不足时, 通过 EdgeBox 方法提取目标候选区域, 利用设计的重检测滤波器, 确定目标的最终位置。在标准数据集 OTB-100 和 TC-128 上的实验结果表明, 本文算法取得了较高的跟踪精度, 在目标发生遮挡、光线变化、出视野等复杂情况时, 算法依旧可以稳健地跟踪。

关键词 机器视觉; 目标跟踪; 相关滤波; 卷积神经网络; 模型更新

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.191502

Multi-Filter Collaborative Tracking Algorithm Based on High-Confidence Updating Strategy

Zhang Chaoyi, Peng Li, Jia Tianhao, Wen Jiwei*

Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education,
School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract A multi-filter collaborative tracking algorithm based on high-confidence updating strategy is proposed. First, the multi-layer convolutional features of the region around the target are extracted using VGG-Net-19, which is a convolutional network architecture, followed by an adaptive feature fusion strategy with the designed deep filter to get the initial position of the target. Meanwhile, a scale filter is constructed to detect the size change of the target. Then, a tracking confidence indicator named primary and secondary peak slope ratio is utilized, which helps to build a high-confidence model updating strategy. Finally, when the confidence is insufficient, the object region proposals are extracted by EdgeBox method, and the final position of the target is determined by the designed re-detection filter. The experimental results on OTB-100 and TC-128 datasets show that the proposed algorithm achieves high tracking precision and also tracks steadily under some complex circumstances, such as occlusion, illumination variation, and out-of-view.

Key words machine vision; object tracking; correlation filter; convolutional neural network; model update

OCIS codes 150.1135; 100.4999; 100.3008

1 引言

目标跟踪是计算机视觉领域的一个重要问题, 在视频监控、人机交互、自动驾驶等领域都有广泛的应用^[1]。视觉跟踪的典型场景是在第一帧中以包围框的形式给定任意一个目标的初始位置和尺寸, 跟踪器需要在接下来的图像序列中预测出该物体的位

置。尽管最近几十年跟踪领域取得了重大的突破, 但受光照变化、遮挡、背景干扰等诸多影响, 目标外观的变化性很大, 因此视觉跟踪仍然是一个极具挑战性的问题。

近年来, 基于相关滤波的跟踪器由于具有较好的准确度和实时性, 吸引了相关学者的广泛关注。Bolme 等^[2]利用图像的灰度特征构建了一个最小输

收稿日期: 2019-04-02; 修回日期: 2019-04-09; 录用日期: 2019-04-18

基金项目: 国家自然科学基金(61873112)、教育部-中国移动科研基金(MCM20170204)

* E-mail: wjw8143@aliyun.com

出平方误差和的跟踪器模型,开创了相关滤波在视觉目标跟踪领域的先河;Henriques等^[3]对循环移位密集采样得到的图像块进行傅里叶变换,在频域内高效地训练滤波器,并使用方向梯度直方图(HOG)特征提升对目标的描述能力,极大地提高了跟踪器的准确性;Danelljan等^[4]将颜色属性特征引入相关滤波框架,使得跟踪器能更好地适应彩色序列;Bertinetto等^[5]在相关滤波框架下将HOG特征和颜色统计特征进行融合互补,进一步提升了跟踪性能。为解决跟踪过程中目标的尺度变化,Li等^[6]结合尺度金字塔与相关滤波,对不同尺度的图像块进行检测,以最大响应值对应的尺度作为当前最佳尺度;Danelljan等在文献[7]中提出一维尺度滤波器,并在文献[8]中通过减少尺度检测数以及对HOG特征进行降维,提高了跟踪速度。

上述跟踪算法都是采用传统的手工特征对目标进行描述,近些年,由卷积神经网络提取的深度特征在很多视觉任务上都取得了不错的表现,在视觉跟踪领域,基于深度特征的跟踪算法^[9-13]也表现出了优秀的性能。Ma等^[9]提出基于分层卷积特征的目标跟踪算法,利用3层卷积特征和由粗糙到细粒的搜索思想来确定目标位置;Qi等^[12]在充分利用多层卷积特征的基础上,采用自适应对冲方法将多个弱跟踪器融合为一个强跟踪器,进一步提升了跟踪精度;Wang等^[13]通过分析卷积网络不同层提取的特征,构建了特征筛选网络和两个互补的预测网络,有效抑制跟踪器的漂移。

模型更新也是跟踪器的一个重要组成部分。大多数现有的跟踪算法采用逐帧更新的方式,以适应目标的外观变化,但是当目标被干扰时,这种更新方式极易导致跟踪漂移。最近,一些学者从响应图的角度考虑模型更新问题,Ma等^[14]利用响应图的最大峰值判别是否更新;Wang等^[15]提出平均峰值相关能量,从响应图的波动程度确定更新策略。

尽管上述算法在跟踪领域已经做出了巨大的改进,但是它们大多针对某一个方面去完善,因此在整体的跟踪成功率和稳健性上仍然有待提高,尤其是在一些复杂场景下,如目标被严重遮挡甚至出视野时,这些算法都极易跟踪失败。为此,本文结合位置预测、尺度估计、模型更新,提出一种基于高置信度更新策略的多滤波器协同跟踪算法。在相关滤波框架下,设计一种多滤波器协同合作的方式以实现目标跟踪,提出一种跟踪置信度指标(主次峰坡度比)用以判断跟踪结果是否可靠,并依据该指标设计了

高置信度下的更新策略,有效减缓了模型退化问题;利用本文算法在OTB-100^[16]和TC-128^[17]这2个大规模数据集上进行测试。通过详细的实验,来验证本文跟踪算法具有较高的跟踪精度与稳健性。

2 基于高置信度更新策略的多滤波器协同跟踪算法

设计3种相关滤波器协同完成跟踪任务,分别从位置计算、尺度估计、目标丢失后的重检测这3个角度考虑,构建不同的滤波器,相互合作。相比之前的跟踪算法,本文算法更加全面,能有效提升整体的跟踪成功率。同时,与传统的相关滤波类方法的模型更新策略相比,本文算法使用更可靠的高置信度更新策略,能减少误差积累,缓解模型退化问题。

2.1 基于深度滤波器的目标初定位

由于不同深度的卷积层对物体的表达能力不一致,为充分描述目标的外观,使用VGG-19深度卷积网络的Conv1-2, Conv2-2, Conv3-4, Conv4-4和Conv5-4这5层来提取目标特征,构建深度滤波器。给定跟踪目标的多通道卷积特征 \mathbf{x} , \mathbf{x} 的大小是 $M \times N \times D$,其中, M , N , D 分别表示宽度、高度和通道数。将目标图像块沿着水平和垂直方向分别进行循环移位,得到 $M \times N$ 个循环移位样本 $\mathbf{x}_{m,n}$,这些训练样本的标签满足高斯分布:

$$y_{m,n} = \exp\left[-\frac{(m-M/2)^2 + (n-N/2)^2}{2\sigma^2}\right], \quad (1)$$

式中, $y_{m,n}$ 为第 m 行第 n 列样本的标签, m 为行数, n 为列数, σ 为高斯核带宽。通过求解最小化问题来获得相关滤波器 \mathbf{W} ,最优滤波器可表示为

$$\mathbf{W}^* = \operatorname{argmin}_{\mathbf{W}} \sum_{m,n} \|\mathbf{W} \cdot \mathbf{x}_{m,n} - y_{m,n}\|^2 + \lambda \|\mathbf{W}\|_2^2, \quad (2)$$

式中, λ 是正则项系数,用来抑制过拟合, $\lambda \geq 0$ 。在每一个独立的特征通道使用快速傅里叶变换得到(2)式的闭式解。第 d 个特征通道上的最优滤波器在频域内可表示为

$$\mathbf{W}^{(d)} = \frac{\mathbf{Y} \odot \overline{\mathbf{X}}^{(d)}}{\sum_{i=1}^D \mathbf{X}^{(i)} \odot \overline{\mathbf{X}}^{(i)} + \lambda}, \quad (3)$$

式中, i 为第 i 个特征通道, \mathbf{Y} , \mathbf{X} , \mathbf{W} 分别为 \mathbf{y} , \mathbf{x} , \mathbf{w} 的离散傅里叶变换, \mathbf{y} 为空域下的样本标签, \mathbf{w} 为空域下的滤波器, $\overline{\mathbf{X}}$ 为 \mathbf{X} 的共轭复数, \odot 为点乘运算。给定目标潜在区域的多通道特征 \mathbf{z} , \mathbf{z} 的大小也是 $M \times N \times D$,其离散傅里叶变换为 \mathbf{Z} ,特征响应可表示为

$$f = \mathcal{F}^{-1} \left(\sum_{d=1}^D \mathbf{W}^{(d)} \odot \overline{\mathbf{Z}}^{(d)} \right), \quad (4)$$

式中, \mathcal{F}^{-1} 是傅里叶逆变换, $\overline{\mathbf{Z}}$ 是 \mathbf{Z} 的共轭复数。利用(4)式求得所用 5 个卷积层的特征响应图 f_l , l 为卷积层索引, $l \in \{1, 2, 3, 4, 5\}$, 通过自适应地融合这 5 层响应图, 以最大响应位置作为目标中心点, 即完成了目标的初定位。

采用的自适应融合策略由 2 部分组成。首先, 因为深层特征可以从高级的语义层面描述目标, 相比浅层特征更具备泛化性和稳健性, 因此赋予深层特征图更大的固定权重, 为

$$\mu_{A,l} = \{2^{-4}, 2^{-3}, 2^{-2}, 2^{-1}, 1\}. \quad (5)$$

由于最大响应值在一定程度上可以反映跟踪置信度, 同时观察到浅层特征响应图的最大响应值比深层的大, 因此采用自适应的方式进行处理:

$$\mu_{B,l} = \frac{1}{\max(f_l)} \bigg/ \left[\sum_{l=1}^5 \frac{1}{\max(f_l)} \right]. \quad (6)$$

对(5)式和(6)式进行融合得到不同层相应的权重, 可表示为

$$\mu_l = \frac{\mu_{A,l} \odot \mu_{B,l}}{\sum_{l=1}^5 \mu_{A,l} \odot \mu_{B,l}}, \quad (7)$$

通过寻找融合特征响应图的最大值所在的位置来确定目标此时的中心位置:

$$\operatorname{argmax}_{x,y} \sum_{l=1}^5 \mu_l \odot f_l(x,y). \quad (8)$$

2.2 基于高置信度的模型更新策略

相关滤波跟踪器通常采用逐帧更新模型的方式, 以适应目标发生的一系列变化, 实际上, 这很容易产生模型退化问题。当某一帧已经跟踪不准确甚至漂移的时候, 尤其是当目标已经被遮挡或发生出视野的情况时, 如果仍然更新模型, 跟踪器就会学习到遮挡物和杂乱背景这些不是目标本身的特征, 误差逐渐累积, 模型被逐渐污染, 等到物体重新出现时, 跟踪器最终无法识别出目标导致跟踪失败。

为解决模型退化问题, 以跟踪结果响应图作为反馈, 确定是否需要在当前帧更新模型。理想的响应图应当是只有一个尖锐的单峰且该峰值较大, 其他部分都比较平滑, 这意味着跟踪器完美地匹配上目标。而当响应图出现多峰, 且波动程度较大时, 跟踪器此时感受到的物体不一定是真实的目标, 很大程度上已经发生了漂移, 此时就应当停止模型的更新, 防止跟踪器被污染。因此本文提出一种跟踪结果置信度指标, 即主次峰坡度比 R_{PSPS} , 可表示为

$$R_{\text{PSPS}} = \frac{y_{\max_P} - y_{\max_S}}{\exp \left[\frac{1}{\sqrt{(x_{\max_{P_x}} - x_{\max_{S_x}})^2 + (y_{\max_{P_y}} - y_{\max_{S_y}})^2}} \right]} \quad \text{s.t.} \quad y_{\max_S} > \frac{y_{\max_P}}{3}, \quad (9)$$

式中: y_{\max_P} 表示最大峰值, 即主峰值; y_{\max_S} 表示第二峰值, 即次峰值。 $(x_{\max_{P_x}}, y_{\max_{P_y}})$ 和 $(x_{\max_{S_x}}, y_{\max_{S_y}})$ 分别表示主峰、次峰所在的横纵坐标。 R_{PSPS} 值结合响应图的最大值与波动程度的关系如图 1 所

示, 当响应图接近理想状态时, 副峰值很小且距离主峰较远, 此时 R_{PSPS} 值较大; 当响应图振荡较大时, 主峰与副峰很接近, 包括峰值上的接近和距离上的接近, 此时 R_{PSPS} 值就会很小。

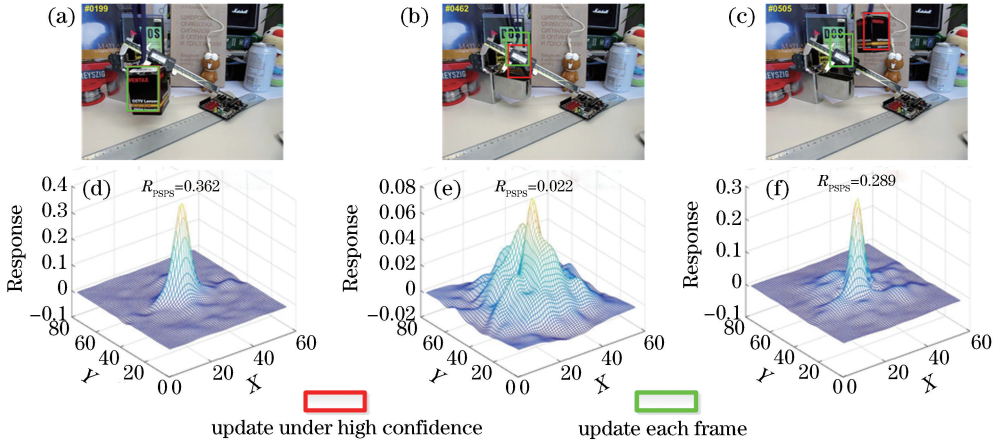


图 1 Box 序列的跟踪结果与响应图。(a)~(c)第 199、462、505 帧图像;(d)~(f)第 199、462、505 帧图像的响应图

Fig. 1 Tracking results and response maps in Box sequence. (a)-(c) 199th, 462nd, and 505th frame images;

(d)-(f) 199th, 462nd, and 505th frame response maps

利用 R_{PSPS} 作为跟踪结果置信度指标,决策是否需要更新模型,采用线性插值方式,同时对滤波器 $\mathbf{W}_t^{(d)}$ 的分子分母进行更新,具体为

$$\begin{cases} \mathbf{A}_t^{(d)} = (1 - \varepsilon\eta)\mathbf{A}_{t-1}^{(d)} + \varepsilon\eta\mathbf{Y} \odot \bar{\mathbf{X}}^{(d)} \\ \mathbf{B}_t^{(d)} = (1 - \varepsilon\eta)\mathbf{B}_{t-1}^{(d)} + \varepsilon\eta \sum_{i=1}^D \mathbf{X}_t^{(i)} \odot \bar{\mathbf{X}}^{(i)} \\ \mathbf{W}_t^{(d)} = \frac{\mathbf{A}_t^{(d)}}{\mathbf{B}_t^{(d)} + \lambda} \\ \varepsilon = \begin{cases} 1, R_{\text{PSPS}} \geq T_{\text{up}} \\ 0, R_{\text{PSPS}} < T_{\text{up}} \end{cases} \end{cases}, (11)$$

式中, $\mathbf{A}_t^{(d)}$ 为第 t 帧滤波器第 d 个特征的分子部分, $\mathbf{B}_t^{(d)}$ 为第 t 帧滤波器第 d 个特征的分母部分, ε 为判断是否需要更新的标志位, η 为更新速率, T_{up} 为决策更新的阈值,上标 (d) 为第 d 个特征通道,下标 t 为第 t 帧。

从图 1 可以看出,使用高 R_{PSPS} 值时更新的策略可以有效缓解模型退化问题。以 Box 序列为例,目标在 461~488 帧期间被严重遮挡, R_{PSPS} 衰减至接近 0,此时选择不更新模型,防止模型被遮挡物污染,当目标重新出现时,跟踪器成功找回目标。图 2 为 2 种更新策略在 Box 序列的中心位置误差。可以看出,如果采用传统的逐帧更新的方式,在发生严重遮挡后,跟踪器随即丢失并且再也无法匹配上目标。

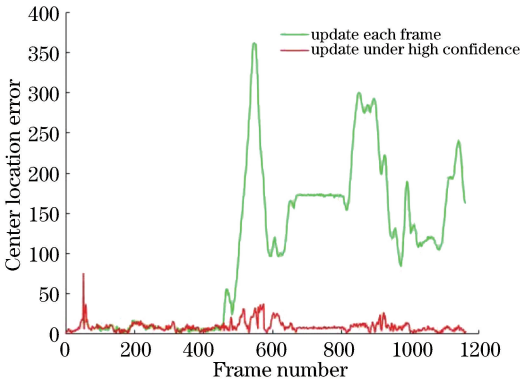


图 2 Box 序列在 2 种更新策略下的中心位置误差
Fig. 2 Center location errors of two updating strategies in Box sequence

另外,对于大多数相关滤波类的由响应图决策跟踪结果的跟踪算法,主次峰坡度比这一置信度指标可以便捷地帮助跟踪器判断当前跟踪结果是否可靠,及时减少误差的积累。

2.3 目标重检测机制

当目标发生遮挡、出视野等情况时,虽然利用高置信度模型更新策略可以有效缓解模型退化问题,但是依然存在跟踪失败的情况,为进一步提升跟踪器的成功率,本文建立低置信度下的目标重检测机

制。该机制由 2 部分组成:1)使用 EdgeBox^[18]在一定搜索范围内生成潜在的目标框;2)构建基于 Hog 特征的重检测滤波器,使用其对潜在的目标框进行评分,当满足一定条件时采用重检测的结果。

2.3.1 EdgeBox 检测器

EdgeBox 方法利用物体的边缘信息可以快速地提取出高质量的目标候选区域,与传统的滑动窗口扫描整幅图像相比,该方法可以消除大量冗余的窗口,筛选出较精确地包含目标的候选框。一些研究人员也将该方法应用在跟踪领域^[19-21]。

2.3.2 重检测滤波器

在深度滤波器得到目标中心后,裁剪出与目标尺寸相同的区域并提取该区域的 HOG 特征,然后使用重检测滤波器计算该区域的响应,设定 T_0 和 T_{re} 这 2 个阈值来判断是否跟踪失败,如果响应的峰值小于 T_0 且响应的 R_{PSPS} 值小于 T_{re} ,那么就认为深度滤波器得到的目标中心不准确,需要通过重检测滤波器再次定位。利用 EdgeBox 方法生成高质量的候选框 B_i ,提取候选框内区域的 HOG 特征,再次利用重检测滤波器计算每一个候选框的响应值 g ,得到目标的最终位置为

$$\begin{aligned} & \operatorname{argmax}_j g_{R_{\text{PSPS}}}(B_j) \\ \text{s.t. } & g_{R_{\text{PSPS}}}(B_j) > 1.5T_{\text{re}} \ \& \ g_{\max}(B_j) > 1.5T_0, \end{aligned} \quad (12)$$

式中, $g_{R_{\text{PSPS}}}(B_j)$ 和 $g_{\max}(B_j)$ 分别为第 j 个候选框的响应 R_{PSPS} 值及其最大值。

2.4 尺度估计

在得到目标的中心位置后,对目标区域的图像块进行缩放,构建图像金字塔,完成多尺度采样。每个图像块的大小为

$$S_{k,t} = a^{(k)}\omega_{t-1} \times a^{(k)}h_{t-1}, \quad (13)$$

式中, $S_{k,t}$ 表示第 t 帧第 k 个图像块的大小, $k \in \left\{ -\left(\frac{n-1}{2}\right), \dots, \left(\frac{n-1}{2}\right) \right\}$, n 表示采样数目, a 表示尺度因子, ω_{t-1} 和 h_{t-1} 分别表示第 $t-1$ 帧目标的宽度和高度。提取不同尺度图像块的 HOG 特征,构建尺度滤波器:

$$\mathbf{H}_s = \frac{\mathbf{G}_s \odot \bar{\mathbf{F}}_s}{\mathbf{F}_s \odot \bar{\mathbf{F}}_s + \lambda_s}, \quad (14)$$

式中, \mathbf{G}_s 是尺度滤波器的标签经傅里叶变换后的频域表示, \mathbf{F}_s 是图像块的 HOG 特征经傅里叶变换后的频域表示, $\bar{\mathbf{F}}_s$ 是 \mathbf{F}_s 的共轭复数, λ_s 是尺度正则项系数。检测当前帧的尺度时,提取不同尺度因子下图像

块的 HOG 特征 F_s , 则不同尺寸因子的响应值为

$$f_s = \mathcal{F}^{-1} \left(\sum_{d=1}^D \overline{F}_s \odot H_s \right). \quad (15)$$

选择最大响应值对应的尺度因子, 完成当前帧目标的尺度估计。同时, 为提高算法速度, 只有当初

定位后得到的响应 R_{PSPS} 值超过尺度更新阈值 T_s 时, 才进行尺度估计。

2.5 算法流程

图 3 为关键环节的具体步骤, 图 4 为整体的框架, 记输入图像序列为 I 。

Input: video sequence I , initial target position and its size (x_1, y_1, w_1, h_1)

Output: estimated target position and its size: (x_t, y_t, w_t, h_t)

1: Repeat

2: crop out the search area in frame t centered at (x_{t-1}, y_{t-1}) with the size of (w_{t-1}, h_{t-1}) , and extract convolutional features to build deep filters, then calculate response map using (4);

3: estimate the new potential position (x_t, y_t) using (8), and calculate the primary R_{PSPS} value using (9) according to the deep filter's response map;

4: if primary R_{PSPS} value is greater than T_s , then estimate the scale of the target using (15);

5: crop out the search window centered at (x_t, y_t) with the size of (w_t, h_t) , and extract

HOG feature to construct re-detect filter, then calculate the secondary R_{PSPS} value

according to this filter's response map;

6: if satisfy the re-detection conditions stated in section 3.3.2, then perform target re-detection using (12) to get final position (x_t, y_t) ;

7: if the primary R_{PSPS} value is greater than T_{upc} , then update deep filters using (10)-(11);

图 3 算法具体步骤

Fig. 3 Specific steps of the proposed algorithm

3 实验结果与分析

3.1 实验平台与参数设置

实验算法采用 Matlab 与 C++ 混合编程, 均在 Windows 10 操作系统下进行。处理器为 Intel i7-8700, 主频是 3.20 GHz, 内存是 16 G, 显卡是 NVIDIA Geforce GTX 1070, 显存是 8 G。算法主要参数: 模型更新率 η 为 0.01, 深度滤波器模型更新的阈值 T_{upc} 和进行尺度估计的阈值 T_s 为 0.075, 进行重检测的阈值 T_0 和 T_{re} 分别为 0.3 和 0.075, 重检测滤波器的更新阈值 T_{upr} 为 0.45。尺度检测参数: 尺度因子 a 为 1.02, 尺度采样数目 n 为 33。EdgeBox 中提取的最大候选框数目在权衡算法速率后选择为 200。

3.2 实验数据与评估方式

为充分验证跟踪算法的精确性与稳健性, 首先

采用 OTB-100 标准数据集中全部 100 个序列进行测试, 这 100 个序列包含快速运动、尺度变化、遮挡、出视野等 11 个跟踪领域常见的难点, 因此能够对跟踪算法进行全面且客观的评价。

实验采用一次通过评估 (OPE) 的方式, 即根据第一帧中对目标的标注位置进行初始化, 然后在该序列中运行算法直至结束。实验采用距离精度和成功率曲线下的面积这 2 个指标来衡量跟踪算法的准确性。

1) 距离精度: 中心位置误差小于 20 pixel 的帧数占视频总帧数的百分比。中心位置误差为算法跟踪的目标中心与实际目标中心的欧氏距离, 可表示为

$$V_{\text{CLE}} = \sqrt{(x_T - x_G)^2 + (y_T - y_G)^2}, \quad (16)$$

式中, (x_T, y_T) 是跟踪算法预测的目标中心, (x_G, y_G) 是真实的目标中心。

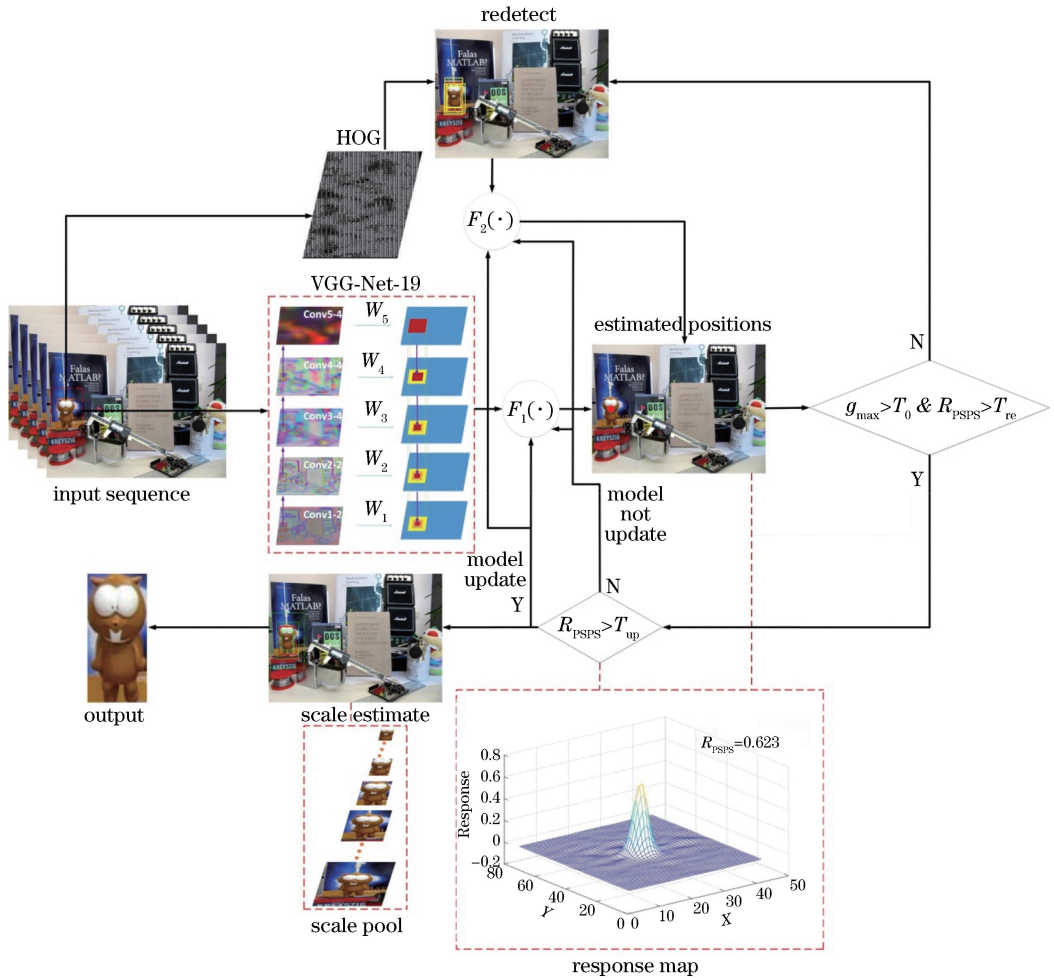


图 4 本文算法的整体框架

Fig. 4 Overall framework of the proposed algorithm

2) 成功率: 跟踪算法预测的目标包围框与真实标注的包围框的交并比(IOU)大于 0.5 的帧数占视频总帧数的百分比。IOU 的计算公式为

$$S = \frac{R_T \cap R_G}{R_T \cup R_G}, \quad (17)$$

式中, R_T 是跟踪算法预测目标的包围框, R_G 是目标真实标注的包围框。本实验利用成功率图曲线下的面积(AUC)作为跟踪算法成功率性能的准则。

3.3 实验设计与分析

为充分验证所提算法的跟踪性能, 将其与 8 种主流算法进行对比, 具体包括核相关滤波跟踪器(KCF), 多尺度跟踪器(DSST), 多特征融合跟踪器(Staple), 空间正则化相关滤波跟踪器^[22](SRDCF), 卷积神经网络-支持向量机跟踪器^[23](CNN-SVM), 分层卷积网络跟踪器(HCF), 对冲多卷积特征跟踪器(HDT)和长短时记忆跟踪器^[24](LCT2)。其中, KCF、DSST、Staple、SRDCF 和 LCT2 是基于传统特征的相关滤波类方法, CNN-

SVM、HCF、HDT 是基于深度特征的相关滤波类方法。

3.3.1 定性分析

图 5 为 9 种算法在 9 组具有各种难点的视频序列上的跟踪结果, 这 9 组序列的属性见表 1。其中, IV 为光照变化, SV 为尺度变化, OCC 为目标被遮挡, DEF 为目标形变, MB 为运动模糊, FM 为快速运动, IPR 为平面内旋转, OPR 为平面外旋转, OV 为目标出视野, BC 为背景扰动, LR 为图像低分辨率。

从目标被遮挡、目标快速运动和光照变化这 3 个方面对本文算法与 8 种对比算法进行分析。

1) 目标被遮挡: 以 Lemming, Human3 和 Girl2 这 3 个长时序为例, 图 5(a)~(c) 为 9 种跟踪算法在这 3 个序列部分帧上的表现。这 3 个序列都存在部分遮挡和完全遮挡的情况, 其中 Lemming, Human3 和 Girl2 序列分别在第 341~359 帧, 第 50~59 帧和第 108~122 帧发生完全遮挡, 当目标



图 5 9 种跟踪算法在 9 组序列下的结果。(a) Lemming; (b) Human3; (c) Girl2; (d) DragonBaby; (e) MotorRolling; (f) Human7; (g) Skiing; (h) Shaking; (i) Tiger2

Fig. 5 Results of nine tracking algorithms for nine sequences. (a) Lemming; (b) Human3; (c) Girl2; (d) DragonBaby; (e) MotorRolling; (f) Human7; (g) Skiing; (h) Shaking; (i) Tiger2

表 1 9 组视频序列及其属性与帧数

Table 1 Attributes and frame numbers of 9 video sequences

Sequence	Attributes	Frame number
DragonBaby	SV, OCC, MB, FM, IPR, OPR, OV	113
Human3	SV, OCC, DEF, OPR, BC	1698
MotorRolling	IV, SV, MB, FM, IPR, BC, LR	164
Skiing	IV, SV, DEF, IPR, OPR	81
Shaking	IV, SV, IPR, OPR, BC	365
Tiger2	IV, OCC, DEF, MB, FM, IPR, OPR, OV	365
Girl2	SV, OCC, DEF, MB, OPR	1500
Human7	IV, SV, OCC, DEF, MB, FM	250
Lemming	IV, SV, OCC, FM, OPR, OV	1336

重新出现时,大部分跟踪器都无法再跟上目标,而本文算法可以在发生严重遮挡后重新找回目标,表现了良好的稳健性和长时跟踪的准确性。抗遮挡性要归功于本文所采用的高置信度模型更新策略和重检

测滤波器,其他算法由于采用逐帧更新的方式,在目标被遮挡时模型会学习到遮挡物的特征,导致误差随时间的推移逐步积累,模型会被遮挡物严重污染,而且除 LCT2 以外,其他算法都没有重检测模块,因此当目标重新出现后,跟踪器无法再匹配上目标,从而产生漂移,而 LCT2 虽然具备重检测模块,但是由于其特征无法区分目标与背景,因此早早地跟丢了目标。另外,在 Girl2 序列中,如果只是采用高置信度模型更新策略,在发生完全遮挡后依旧跟踪失败,当加上重检测滤波器后才得以找回目标。

2) 目标快速运动:以 DragonBaby, MotorRolling 和 Human7 这 3 个序列为例,图 5(d)~(f)为 9 种跟踪算法在这 3 个序列部分帧上的表现。这 3 个序列中都包含了剧烈的快速运动,相邻帧的目标中心位置可能会相距比较大,同时又都具备运动模糊,进一步加大了跟踪难度。可以看出本文算法在这些困难环境下都取得了较好的结

果,同时对于尺度的变化也能及时做出调整。其他算法中基于深度特征的跟踪器(HDT、CNN-SVM、HCF)虽然可以准确计算出目标中心位置,但在尺度方面都没有做出应对,而基于传统特征的跟踪器由于特征对外界变化比较敏感,因此极易产生漂移现象,导致跟踪失败。

3)光照变化:以 Skiing, Shaking 和 Tiger2 这 3 个序列为例,图 5(g)~(i)为 9 种跟踪算法在这 3 个序列部分帧上的表现。Skiing 序列中运动员从第 5~30 帧完成了 360°的空翻,所有的传统特征类方法全部失败,只有基于深度特征的算法可以跟上目标。Shaking 序列中吉他手周边的环境光线变化非常剧烈,在第 61 帧发生强光照时,Staple, KCF 和 DSST 算法立刻产生漂移并再也没有找回目标,而本文算法自始至终都没有受外界光照影响,可以紧跟目标。Tiger2 序列中目标在灯光照射下明暗不一并伴随部分遮挡现象,本文算法凭借强大的深度特征和灵活的模型更新策略依旧可以准确地跟踪目标。可以看出,深度特征具备很强的语义信息,对颜

色变化、形变、光照变化都具备较强的稳健性,因此基于深度特征的跟踪器相对于传统特征的跟踪器而言,抗干扰能力更强。

3.3.2 定量分析

图 6 为本文算法对比其他 8 种算法的整体跟踪性能,包括距离精度图和成功率图。在这 2 个评价指标上,本文算法均排名第一。距离精度图中本文算法的得分为 0.862,领先第二、第三和第四的 HDT (0.848)、HCF (0.842) 和 CNN-SVM (0.814) 各 1.4%, 2%, 4.8%。值得一提的是,本文算法与这 3 种算法都是基于深度特征的算法,而基于深度特征的跟踪器在中心距离精度上明显优于基于传统特征的跟踪器。成功率图中本文算法得分为 0.625,领先第二、第三和第四的 SRDCF (0.598)、Staple (0.578) 和 HCF (0.565) 各 2.7%, 4.7%, 6%。尤其是与相似的 HCF 跟踪器相比,本文算法由于结合尺度滤波器、重检测滤波器及更稳健的模型更新方式,同时扩充使用 5 层卷积特征,因此跟踪成功率可得到较大的提升。

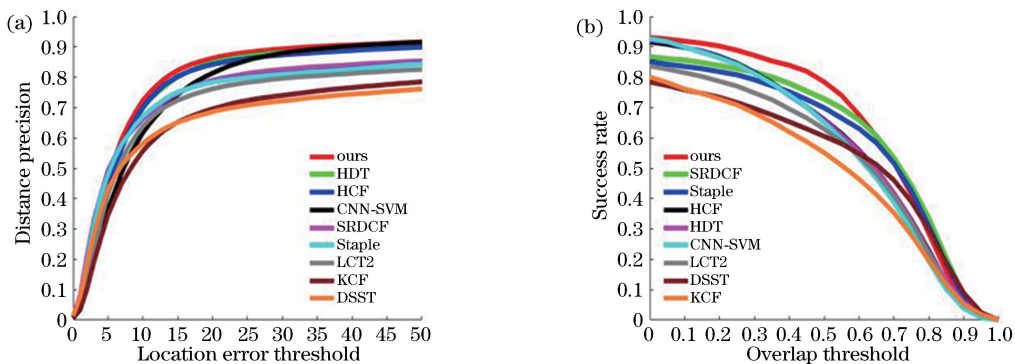


图 6 9 种算法在 OTB-100 序列上的总体跟踪性能。(a)距离精度曲线;(b)成功率曲线

Fig. 6 Overall performance of 9 tracking algorithms on OTB-100. (a) Distance precision curve; (b) success rate curve

表 2 和表 3 分别为本文算法对比其他 8 种算法在 11 种不同属性的距离精度和 AUC 值。在距离精度上,本文算法在 FM, BC, MB, IV, OCC, OPR, OV 和 SV 这 8 个属性上均为最优,尤其是在背景干扰、遮挡和出视野这 3 项属性上分别超过第二名跟踪器 3.8%, 3.7% 和 7.4%。在成功率 AUC 指标上,本文算法除在 LR 这个属性上落后 SRDCF 处于次优,在其他 10 个属性上均为最优。尤其是在背景干扰、运动模糊和出视野这 3 项属性上分别超过第二名的跟踪器 5.2%, 5.4% 和 8.4%。

3.3.3 算法结构分析

为说明本文算法的有效性,设计了组内对照实验。将本文算法拆分成 3 个部分:没有重检测

滤波器同时采用逐帧更新的方式(Alg1),没有重检测滤波器但是采用高置信度更新的策略(Alg2),有重检测滤波器但采用逐帧更新的方式(Alg3)。用这 3 种跟踪方法与本文算法在 100 个序列上进行测试,图 7 为总体的跟踪性能。可以看出,本文算法由于结合了重检测滤波器与高置信度更新策略,总体上性能均达到最优。其中,只有在高置信度下更新模型的策略取得了与本文算法比较接近的效果,这充分验证主次峰坡度比作为置信度指标的可靠性。而本文算法又在此基础上增加了重检测模块,因此能够在目标被严重遮挡甚至出视野的情况下依旧能够成功找回目标,进一步提升了总体的成功率。

表 2 9 种算法在 11 种属性上的距离精度值

Table 2 Distance precision scores of nine algorithms in terms of 11 attributes

Algorithm	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
Ours	0.822	0.882	0.833	0.821	0.866	0.855	0.870	0.815	0.850	0.751	0.830
HDT	0.817	0.844	0.789	0.817	0.820	0.844	0.887	0.774	0.805	0.663	0.808
HCF	0.815	0.843	0.804	0.791	0.831	0.864	0.847	0.778	0.816	0.677	0.799
CNN-SVM	0.747	0.776	0.751	0.793	0.795	0.813	0.925	0.730	0.798	0.650	0.787
SRDCF	0.769	0.775	0.767	0.734	0.792	0.745	0.765	0.735	0.742	0.597	0.745
Staple	0.710	0.749	0.699	0.751	0.782	0.768	0.695	0.728	0.738	0.668	0.727
LCT2	0.681	0.734	0.669	0.689	0.746	0.781	0.699	0.682	0.746	0.592	0.681
KCF	0.621	0.713	0.601	0.617	0.719	0.701	0.671	0.630	0.677	0.501	0.633
DSST	0.571	0.704	0.568	0.550	0.722	0.698	0.684	0.603	0.655	0.478	0.649

表 3 9 种算法在 11 种属性上的 AUC 值

Table 3 AUC scores of nine algorithms in terms of 11 attributes

Algorithm	FM	BC	MB	DEF	IV	IPR	LR	OCC	OPR	OV	SV
Ours	0.622	0.637	0.648	0.583	0.638	0.597	0.489	0.605	0.601	0.572	0.583
HDT	0.568	0.578	0.574	0.543	0.535	0.555	0.401	0.528	0.533	0.472	0.486
HCF	0.570	0.585	0.585	0.530	0.549	0.566	0.388	0.532	0.540	0.474	0.485
CNN-SVM	0.546	0.548	0.578	0.547	0.537	0.548	0.379	0.515	0.548	0.488	0.490
SRDCF	0.597	0.583	0.594	0.544	0.613	0.544	0.514	0.559	0.550	0.460	0.561
Staple	0.541	0.561	0.540	0.550	0.593	0.548	0.399	0.542	0.533	0.476	0.520
LCT2	0.507	0.528	0.516	0.482	0.517	0.529	0.330	0.477	0.505	0.446	0.430
KCF	0.459	0.498	0.459	0.436	0.479	0.469	0.290	0.443	0.453	0.394	0.394
DSST	0.458	0.523	0.469	0.427	0.559	0.505	0.382	0.460	0.477	0.386	0.475

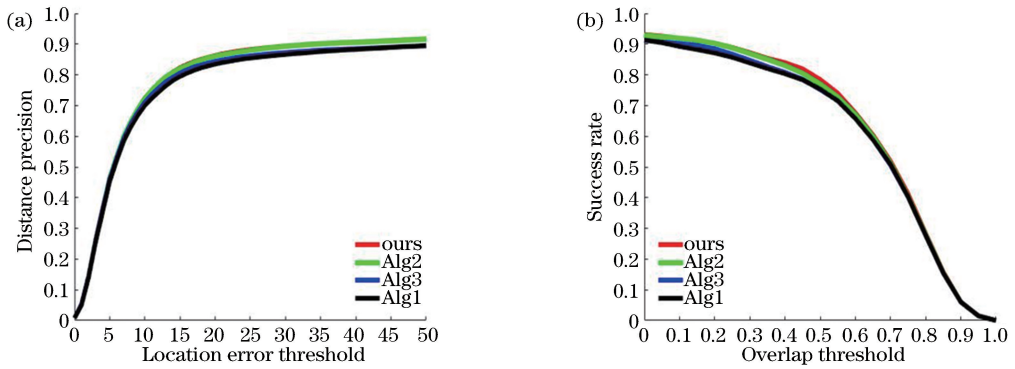


图 7 OTB-100 上的算法结构总体分析。(a)距离精度曲线;(b)成功率曲线

Fig. 7 Overall analysis of algorithm structure on OTB-100. (a) Distance precision curve; (b) success rate curve

3.3.4 置信度参数的设定

本文算法采用高置信度的模型更新策略,因此深度滤波器的更新阈值 T_{upc} 对跟踪成功率有着重要的意义。如果 T_{upc} 较大,就意味着更新条件较严格,模型会学习到更加可靠的目标特征,但也会在一定程度上忽略目标细微的变化;如果 T_{upc} 较小,那么模型的学习又相对容易受到干扰物的影响导致误差积累。因此,为找到最优的更新阈值,设定 T_{upc} 的取值集合 $S = \{0.050, 0.075, 0.100, 0.125\}$, 在 OTB-100 数据集上进行逐一测试,结果如图 8 所示。记 T_{upc} 取值分别为 0.050, 0.075, 0.100, 0.125 时算法名称

为 T1,ours,T2,T3。可以看出,当 T_{upc} 取值为 0.075 时,算法的跟踪效果最好。

3.3.5 算法时间分析

表 4 列出了本文算法与其他 8 种对比算法在 OTB-100 序列上的平均速度,其中,“—”表示没有给出算法的平均速率。可以看出,与使用传统的手工特征的跟踪器(表中后 5 种算法)相比,基于深度特征的跟踪器(表中前 4 种算法)虽然在跟踪成功率上表现得更好,但是相对较慢。本文跟踪算法相对 HCF 算法的跟踪成功率有较大提高,但是由于引入尺度估计和重检测机制,即使通过高置信度更新策略

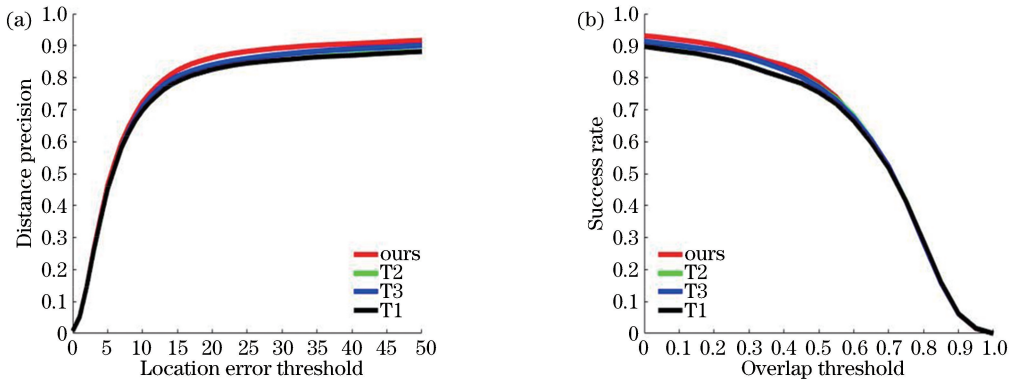


图 8 本文算法在 OTB-100 序列上不同的更新阈值参数下的结果。(a)距离精度曲线;(b)成功率曲线

Fig. 8 Overall performance of our algorithm under different updating parameters on OTB-100. (a) Distance precision curve; (b) success rate curve

表 4 9 种跟踪算法的平均速度对比

Table 4 Comparison of average tracking speeds of 9 tracking algorithms

frame/s

Tracker	Ours	HDT	HCF	CNN-SVM	SRDCF	Staple	LCT2	KCF	DSST
Average tracking speed	7.9	9.1	13.9	—	4.3	56.3	28.1	317	46.4

可以起到一定的提速作用,但是总体上速度还是稍慢。

3.3.6 TC-128 数据集结果评估

为进一步验证本文算法的稳健性,将本文算法与其他算法在 TC-128 数据集上进行测试,评价标准与 OTB-100 一致。图 9 为最终的结果。

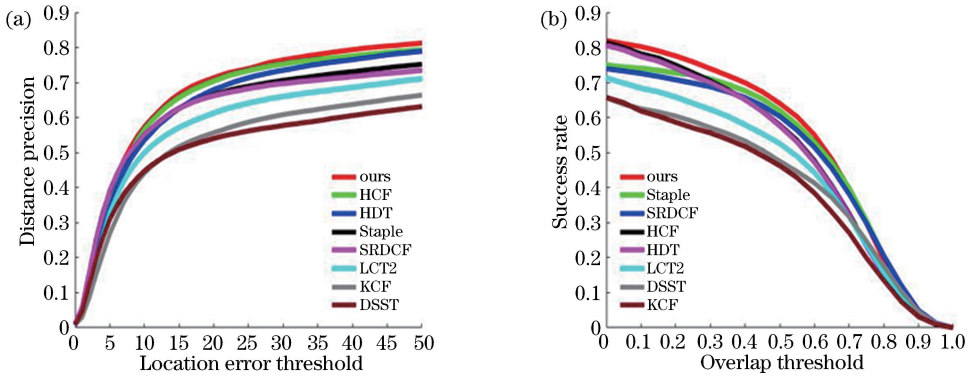


图 9 9 种算法在 TC-128 序列上的总体跟踪性能。(a)距离精度曲线;(b)成功率曲线

Fig. 9 Overall performance of 9 tracking algorithms on TC-128. (a) Distance precision curve; (b) success rate curve

4 结 论

提出一种基于高置信度模型更新策略的多滤波器协同跟踪算法。算法构建 3 种相关滤波器协同完成跟踪任务。利用多层卷积特征构建深度滤波器,以自适应的特征融合策略进行目标初定位;为解决目标的尺度变化,设计尺度滤波器进行准确的尺度估计;提出一种跟踪置信度指标:主次峰坡度比,用以确定跟踪结果是否可靠,从而决策是否需要更新模型,缓解模型退化问题;设计重检测滤波器,当跟踪结果不可靠时,进行目标重定位。在标准数据集

可以看出,本文算法在 TC-128 序列上仍然表现最优。另外,本文算法在 OTB-100 和 TC-128 这 2 个数据集上使用相同的参数,因此进一步验证本文算法具备较强的稳健性,在不同的环境下都能取得较好的结果。

OTB-100 和 TC-128 上进行详实的算法验证,实验结果表明本文算法具备较高的精度与稳健性,尤其是在目标被遮挡、出视野等长时跟踪场景中普遍存在的情况下,依旧可以稳定地跟踪。此外,虽然深度特征对于目标的描述能力与稳健性更强,但是基于深度特征的跟踪算法在速度方面比较慢,因此,未来的研究重点是在保持较高精度的同时提高跟踪器的实时性。

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] 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, 2010: 2544-2550.
- [3] 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.
- [4] 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 (CVPR), June 23-28, 2014, Columbus, OH, USA. New York: IEEE, 2014: 1090-1097.
- [5] 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, 2016: 1401-1409.
- [6] 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. Lecture notes in computer science. Cham: Springer, 2015, 8926: 254-265.
- [7] Danelljan M, Häger G, Khan F S, *et al.* Accurate scale estimation for robust visual tracking [C] // British Machine Vision Conference 2014, September 1-5, 2014, Nottingham, British. Durham, England: BMVA Press, 2014.
- [8] Danelljan M, Hager G, Khan F S, *et al.* Discriminative scale space tracking [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(8): 1561-1575.
- [9] 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, 2015: 3074-3082.
- [10] Wang X, Hou Z Q, Yu W S, *et al.* Target scale adaptive robust tracking based on fusion of multilayer convolutional features [J]. Acta Optica Sinica, 2017, 37(11): 1115005.
- 王鑫, 侯志强, 余旺盛, 等. 基于多层卷积特征融合的目标尺度自适应稳健跟踪 [J]. 光学学报, 2017, 37(11): 1115005.
- [11] Mao N, Yang D D, Yang F C, *et al.* Adaptive object tracking based on hierarchical convolution features [J]. Laser & Optoelectronics Progress, 2016, 53(12): 121502.
- 毛宁, 杨德东, 杨福才, 等. 基于分层卷积特征的自适应目标跟踪 [J]. 激光与光电子学进展, 2016, 53(12): 121502.
- [12] 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, Las Vegas, NV, USA. New York: IEEE, 2016: 4303-4311.
- [13] 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, 2015: 3119-3127.
- [14] 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, 2015: 5388-5396.
- [15] Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 4021-4029.
- [16] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [17] Liang P P, Blasch E, Ling H B. Encoding color information for visual tracking: algorithms and benchmark [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5630-5644.
- [18] Zitnick C L, Dollár P. Edge boxes: locating object proposals from edges [M] // Fleet D, Pajdla T, Schiele B, *et al.* Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8693: 391-405.
- [19] Zhu G, Porikli F, Li H D. Beyond local search: tracking objects everywhere with instance-specific proposals [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York:

- IEEE, 2016: 943-951.
- [20] Ge B Y, Zuo X Z, Hu Y J. Long-term object tracking based on feature fusion [J]. *Acta Optica Sinica*, 2018, 38(11): 1115002.
葛宝义, 左宪章, 胡永江. 基于特征融合的长时目标跟踪算法[J]. *光学学报*, 2018, 38(11): 1115002.
- [21] 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*, 2018.
- [22] Danelljan M, Hager 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, 2015: 4310-4318.
- [23] Hong S, You T, Kwak S, *et al.* Online tracking by learning discriminative saliency map with convolutional neural network [C] // Proceedings of the 32nd International Conference on International Conference on Machine Learning, July 6-11, 2015, Lille, France. New York: ACM, 2015, 37: 597-606.
- [24] Ma C, Huang J B, Yang X K, *et al.* Adaptive correlation filters with long-term and short-term memory for object tracking[J]. *International Journal of Computer Vision*, 2018, 126(8): 771-796.