

结合目标检测的多尺度相关滤波视觉跟踪算法

王红雨*, 汪梁, 尹午荣, 胡江颢, 乔文超

上海交通大学电子信息与电气工程学院, 上海 200240

摘要 为满足视觉跟踪算法对跟踪精度与跟踪速度的要求, 提出一种结合目标检测的多尺度相关滤波视觉跟踪算法。所提算法基于深度学习的目标检测算法找出图像中目标的位置和尺寸, 利用相关滤波算法对所给出的目标特征进行视觉跟踪, 并在多个尺度中搜索最优响应; 当检测到相关滤波响应值异常时, 停止对模型更新; 当连续数帧响应值异常时, 则在全图范围内搜索目标位置和尺寸。所提算法通过对跟踪状态进行评估和模型更新率自适应调整, 解决了传统相关滤波类算法跟踪误差随时间积累的问题, 且具有较大的跟踪速度和较高的精度。结果表明: 在 Matlab 平台下, 所提算法的平均定位精度为 0.593, 平均交叠率精度为 0.784, 帧率为 65.3 frame/s。

关键词 机器视觉; 深度学习; 目标检测; 相关滤波; 视觉跟踪; 响应估计

中图分类号 TP391.41

文献标识码 A

doi: 10.3788/AOS201939.0115004

Multi-Scale Correlation Filtering Visual Tracking Algorithm Combined with Target Detection

Wang Hongyu*, Wang Liang, Yin Wurong, Hu Jianghao, Qiao Wenchao

School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

Abstract In order to satisfy the requirements of visual tracking algorithm on tracking accuracy and speed, a multi-scale correlation filtering visual tracking algorithm combined with target detection is proposed. The proposed algorithm is first used to find the target location and size in the image by the target detection algorithm based on depth learning. The correlation filtering algorithm is then applied to the visual tracking of the given target features and the multi-scale search of the optimal response. When the correlation filtering response appears abnormal, the model stops updating. When the response value of several frames continues to be abnormal, the search of target location and size is then made in the whole image. By the evaluation of tracking states and the adaptive adjustment of model updating rate, the proposed algorithm solves the problem of tracking error accumulation over time in the traditional correlation filter algorithm, and possesses high tracking speed and high precision. The results show that as for the proposed algorithm on the Matlab platform, the average positioning precision is 0.593, the average overlap precision is 0.784, and the frame rate is 65.3 frame/s.

Key words machine vision; deep learning; target detection; correlation filtering; visual tracking; response estimation

OCIS codes 150.0155; 150.1135; 100.4999; 110.1085

1 引 言

目标跟踪是计算机视觉领域中的关键技术问题, 已广泛应用于视觉导航、人机交互、无人驾驶、智能监控等领域。但是, 现实中的目标通常面临着外形变化、尺度变换、光照变化、动态模糊、目标遮挡等问题, 因此实现精确的目标跟踪仍面临着诸多挑战。

目前跟踪算法大致分为生成类模型^[1]和判别类模型^[2]。生成类模型的跟踪算法可对目标区域进行建模, 并在下一帧图像寻找最相似区域的位置, 代表性的算法有 meanshift 算法^[3]和 camshift 算法等^[4]。判别类模型的跟踪算法是将目标跟踪问题转化为二分类问题, 并将目标区域和背景区域分别划分为正样本和负样本, 训练分类器, 在下一帧图像中评估和

收稿日期: 2018-07-06; 修回日期: 2018-08-10; 录用日期: 2018-09-05

基金项目: 国家自然科学基金(61471237)

* E-mail: redrain@sjtu.edu.cn

寻找最优区域,代表性的算法有跟踪-学习-检测算法(TLD)^[5]、核化跟踪检测循环结构(CSK)算法^[6]、核自相关滤波(KCF)算法^[7]、平方误差滤波器最小输出和算法^[8]等。近些年,基于相关滤波(CF)的跟踪方法以其优异的准确率和实时性能成为了研究热点。相关滤波算法通过学习目标区域的相关滤波模型,在下一帧图像中寻找预测分布的最大响应值作为当前的预测位置。Henriques等^[6]提出CSK,通过对目标区域循环移位生成训练样本,将训练样本用于训练分类器。由于分类过程利用了循环矩阵在傅里叶空间可对角化的性质,相关滤波算法将矩阵运算转化为频域的点乘运算,大大降低了运算量,提高了算法的实时性。在CSK算法的基础上,Henriques等^[7]将方向梯度直方图^[9](HOG)特征引入相关滤波,提出了KCF算法,将CSK算法的适用范围从灰度图升级为多通道有色图。但是,目前基于传统相关滤波的算法仍然无法处理目标尺度变换的问题。Danelljan等^[10]引入尺度相关估计,提出了判别式尺度空间跟踪(DSST)算法,该算法可根据被跟踪目标自适应调整边界框大小。但是,基于相关滤波的跟踪算法均需要预先提供目标初始位置和跟踪信息,而且相关滤波类的跟踪算法误差会随时间积累,从而导致跟踪失败。此外,相关滤波类算法不能很好地学习和记忆动态变化目标,这限制了目标跟踪算法在现实中的应用。近几年,国内关于跟踪算法的研究取得较大突破。Wang等^[11]提出平均峰值相关能量(APCE),用于反映跟踪置信度水平和波动程度;沈秋等^[12]提出自适应特征选择的多尺度相关滤波跟踪算法,进一步提升算法的跟踪精度和稳健性;徐玉龙等^[13]针对多尺度问题,提出了一定的改进措施。

实时目标检测算法可通过检测每帧图像目标的大小及位置实现目标跟踪,但是受限于算法的复杂度和硬件资源,很难做到实时、高速。基于传统机器学习,Dalal等^[9]提出HOG+SVM(Support Vector Machine)的行人检测算法,实现了对行人的实时目标检测和跟踪;近些年,卷积神经网络(CNN)凭借高精度及多类别检测成为研究热点。Girshick等^[14-16]提出区域CNN(R-CNN)系列目标检测算法,将检测过程划分为候选区域生成与分类提取两个过程,对目标检测精度和速度进行优化;之后的YOLOv2^[17](You Only Look Once)和SSD^[18](Single Shot MultiBox Detector)算法在产生候选区域的同时进行分类回归,较RCNN类算法的检测

速度有所提升,但是精度略低。

针对动态目标的尺度变化问题以及当前跟踪算法需要预先提供目标初始位置的问题,本文引入SSD目标检测算法对目标进行识别和定位;在获得目标初始位置及大小信息的基础上,引入相关滤波算法,对目标进行跟踪;借鉴DSST算法,引入尺度估计滤波器,解决目标多尺度变化的问题;引入对相关滤波响应值的判别比较,当响应值异常时,停止对模型进行更新;当连续数帧响应值异常时,再次调用目标检测算法,对被跟踪目标重新进行初始化,形成检测-跟踪-检测的闭环视觉跟踪过程。相比于实时目标检测算法,所提算法在检测速度上有所提升,也解决了传统相关滤波跟踪算法误差随时间积累的问题。实验结果表明,所提算法的平均定位精度为0.593,平均交叠率精度(OP)为0.784,帧率为65.3 frame/s。

2 SSD目标检测算法

2.1 模型结构

SSD算法在VGG16^[19]的基础上,去掉最后三层全连接层,添加了10层卷积层(共23层),其结构如图1所示,图中Conv代表卷积层。SSD算法在不同尺度的特征图上采用卷积核输出预测类别和位置偏移量。SSD算法对每层特征图生成系列网格,并在不同网格生成不同宽高比的默认框,设预测过程使用了 m 层特征图,则第 k 个特征图的默认框计算式为

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m - 1}(k - 1), \quad k \in \{1, 2, \dots, m\}, \quad (1)$$

式中: s_{\min} 和 s_{\max} 分别代表特征图中默认框占输入图像的最小、最大比例, $s_{\min} = 0.2, s_{\max} = 0.9$ 。各个默认框宽度和高度的计算公式分别为

$$w_k^a = s_k \sqrt{a_r}, \quad (2)$$

$$h_k^a = s_k / \sqrt{a_r}, \quad (3)$$

式中: a_r 为每个默认框的高宽比, $a_r = \{1, 2, 1/2, 3, 1/3\}$,当高宽比为1时,再增加一种尺度比例 $s'_k = \sqrt{s_k s_{k+1}}$ 的默认框,故对于每层特征图上的各个网格,都对应6种不同比例的默认框。多种尺度的默认框在各层特征图中基本能覆盖各种形状和尺寸的目标。

2.2 损失函数

SSD损失函数由位置损失 $L_{\text{loc}}(x, l, g)$ 和分类损失 $L_{\text{conf}}(x, c)$ 两部分组成,其表达式为

$$L(x, c, l, g) = \frac{1}{N} [L_{\text{conf}}(x, c) + aL_{\text{loc}}(x, l, g)], \quad (4)$$

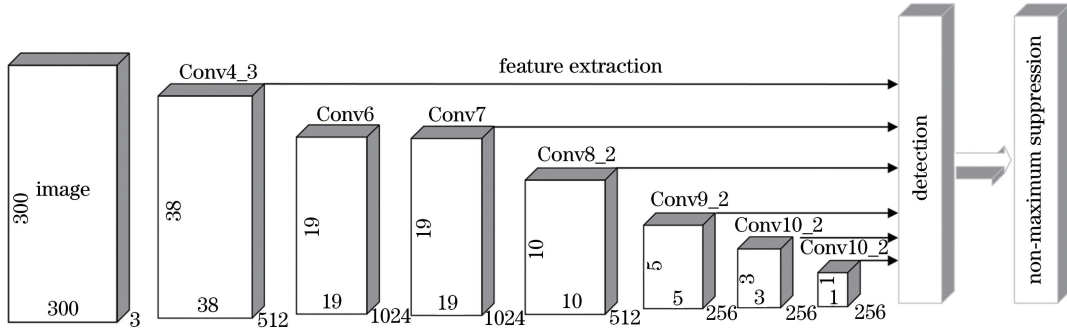


图 1 SSD 网络结构图

Fig. 1 Structural diagram of SSD network

式中： a 为位置损失权重项，通常设为 1； N 为匹配默认框的数量； l 为预测框位置； g 为真实框位置； c 为属于该类别的置信度； x 为与各个类别的匹配结果。通过训练，逐步提高位置预测和分类预测的准确度，从而实现未知样本良好的预测能力。

3 相关滤波算法

相关滤波算法属于判别类视觉跟踪算法，该算法可对目标区域循环移位生成大量样本，从这些样本中学习得到相关滤波分类器；在下一帧图像预测区域采样并进行循环移位操作，生成大量预测样本，计算各个样本的最大响应值，并将拥有最大响应值的预测样本作为下一帧图像中目标的位置。算法中利用循环矩阵生成训练样本，并利用循环矩阵在岭回归中的特殊性质及快速傅里叶变换对算法进行加速。

代价函数可表示为

$$\min_w \sum_i [f(x_i) - y_i]^2 + \lambda \|w\|^2, \quad (5)$$

式中： w 为权重系数， λ 为正则化系数， x_i 为样本， y_i 为回归目标， $f(x_i)$ 为分类器输出。相关滤波算法通过岭回归求解(5)式，得到 w 的闭式解为

$$w = (X^H X + \lambda I)^{-1} X^H y, \quad (6)$$

式中： I 为单位矩阵。使 x_i 和 y_i 的平方误差最小，其中引入 λ 可提高分类器泛化能力，控制过拟合。(6)式中 X 为输入样本，将 X 看作循环矩阵的元素，利用离散傅里叶矩阵 F 对循环矩阵进行对角化，即 $X = F \times \text{diag}(\hat{x}) \times F^H$ ，可以得到

$$\hat{w} = \frac{\hat{x} \odot \hat{y}}{\hat{x} \odot \hat{x}^* + \lambda I}, \quad (7)$$

式中： \hat{w} 、 \hat{x} 、 \hat{y} 分别代表 w 、 x 、 y 的傅里叶变换形式； \odot 代表元素点乘。使用向量点积代替矩阵运算及求逆运算，提高了计算速度。

引入高斯核 $K^{xx} = k(x, x)$ 作为核函数，高斯核的计算公式为

$$K^{xx} = k(x_i, x_j) = \exp\left\{-\frac{1}{\sigma^2} \left\{ \|x\|^2 + \|x'\|^2 - 2\mathcal{F}^{-1}[\mathcal{F}^*(x) \odot \mathcal{F}(x')] \right\}\right\}, \quad (8)$$

式中： σ 为高斯核带宽因子； x' 为 x 变量的转置； \mathcal{F}^{-1} 表示离散傅里叶逆变换， $\mathcal{F}(\cdot)$ 表示离散傅里叶变换。

现实中的样本往往会出现线性不可分的情况，引入非线性函数 $\psi(x)$ 对样本进行重映射，使得样本在高维空间线性可分，在高维空间中， w 的求解可转化为一个线性问题，即 $w = \sum_i a_i \psi(x)$ ， a_i 为 x_i 经过非线性映射的系数， $\psi(x)$ 为对 x_i 的一个非线性映射。任意两个 x 和 x^* 样本在特征空间的相似程度为 $\psi^T(x) \psi(x^*)$ ，记 $\psi^T(x) \psi(x^*) = k(x, x^*) = K^{xx}$ ，则目标函数 $f(z)$ 可表示为

$$f(z) = w^T \psi(z) = \sum_i a_i \psi(x) \psi(z) = \sum_i a_i k(x, z) = a^T k(z). \quad (9)$$

式中： $f(z)$ 为关于 $k(z)$ 的线性函数，采用岭回归的方式对 a 进行求解，可得

$$a = (K^{xx} + \lambda I)^{-1} y, \quad (10)$$

a 经过傅里叶变换得到

$$\hat{a} = \frac{\hat{y}}{K^{xx} + \lambda}, \quad (11)$$

对(9)式两边做傅里叶变换，得到目标函数的傅里叶形式为

$$\hat{f}(z) = \hat{k}(z) \odot \hat{a}, \quad (12)$$

式中 $\hat{k}(z)$ 为 $k(z)$ 的傅里叶变换。

3.1 自适应尺度估计

针对目标出现尺度变化的问题，采用 DSST 算

法,并引入位置相关滤波器和尺度相关滤波器。利用位置相关滤波器估计目标的二维位置信息,利用尺度相关滤波器估计目标的一维尺度信息。该算法估计目标尺度的具体步骤为:

1) 利用位置相关滤波器得到目标预测位置 (x, y) ;

2) 以预测位置为中心,计算 33 个不同尺度的候选区域,利用尺度相关滤波器匹配最佳尺度,候选

区域的尺度选取方式为

$$S_i = a_i w' \times a_i h, \quad (13)$$

式中: $i \in \{-(n-1)/2, \dots, (n-1)/2\}$, $n=33$ 代表 33 个不同尺度; w' 为前一帧目标框的宽度; h 为前一帧目标的高度; $a=1.02$ 代表尺度因子。引入尺度相关滤波器的效果如图 2 所示,可以看出,本文算法能根据目标尺度信息自适应调整边界框大小。

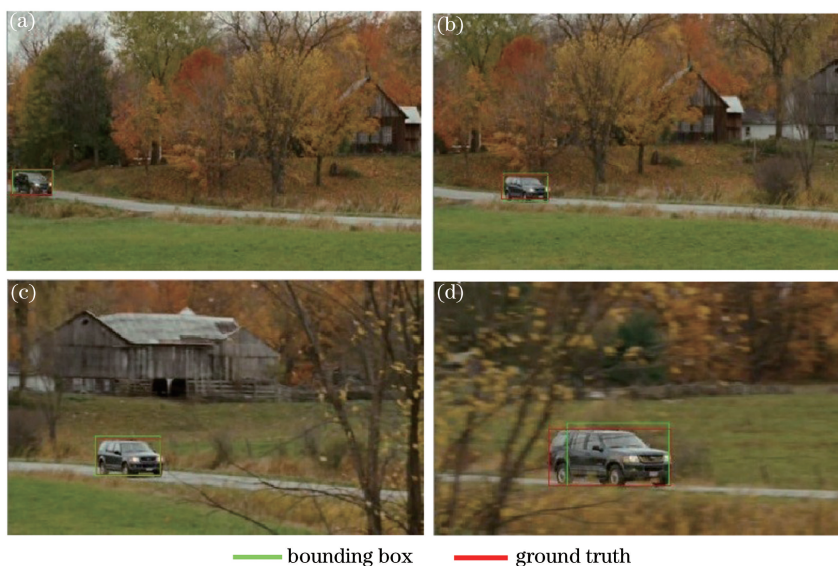


图 2 引入尺度变换的 Carscale 序列

Fig. 2 Carscale sequence with scale transformation

3.2 目标丢失检测

在实际目标跟踪过程中,由于长时间跟踪误差积累及环境复杂等因素的影响,难免出现跟踪失败的问题,当跟踪失败时,需中止对错误目标的跟踪,

并重新对目标的位置和大小信息进行估计。因此,提出一种对跟踪状态的检测和判断机制,用于对当前的跟踪目标置信度进行判断。首先对跟踪成功与跟踪失败的图像序列展开分析,结果如图 3 所示。

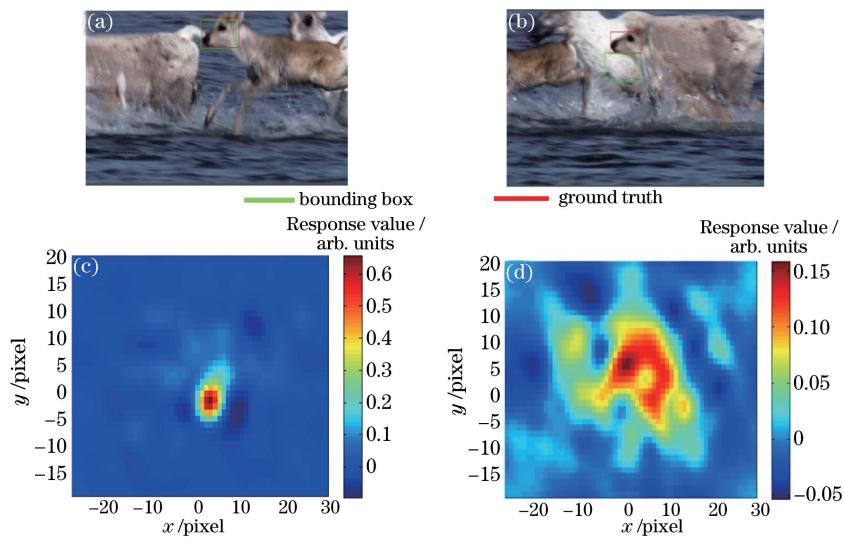


图 3 跟踪序列与相关滤波响应图。(a)跟踪成功序列;(b)跟踪失败序列;(c)跟踪成功序列的相关滤波响应图;(d)跟踪失败序列的相关滤波响应图

Fig. 3 Tracking sequence and correlation filtering response. (a) Tracking success sequence; (b) tracking failure sequence; (c) correlation filtering response of tracking success sequence; (d) correlation filtering response of tracking failure sequence

图 3(c)、(d)为相关滤波响应图,分别对应图 3(a)、(b),由图 3 可知,相关滤波响应值的分布情况与跟踪结果存在关联。当峰值点集中,且峰值较大时,代表跟踪精度较高;当峰值较小,且响应值分散时,则跟踪很可能出现失败。为保证跟踪速度,需要简单高效的判别机制对跟踪结果进

行判断。一种简单的方法是计算响应图的最大响应值 F_{\max} 。一般而言,响应值越大越好,当响应值低于一定阈值时,即判定为跟踪失败,但是此判据并不可靠,当目标出现相似物遮挡时,依旧可以有较高的响应值,如图 4 所示,因此只用该判据依旧容易发生跟踪漂移情况。

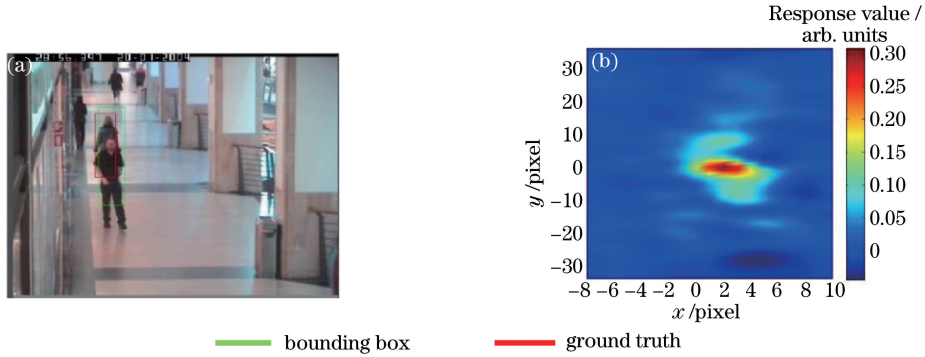


图 4 (a) 目标遮挡的图像; (b) 目标遮挡的相关滤波响应图

Fig. 4 (a) Concealed target; (b) correlation filtering response of concealed target

这里引入 APCE^[11] 指标,该指标可反映响应图峰值水平和波动程度,其计算方式为

$$A_{PCE} = \frac{|F_{\max} - F_{\min}|^2}{\text{mean} \left[\sum_{W,H} (F_{W,H} - F_{\min})^2 \right]}, \quad (14)$$

式中: F_{\max} 、 F_{\min} 分别为响应图的最大、最小值; W 和 H 分别为响应图的宽度和高度; $\text{mean}[\cdot]$ 代表取平均运算。APCE 可用于表示响应图的波动情况和置信程度。峰值越尖锐,起伏越小,则 APCE 值越大。当目标出现遮挡或错误跟踪时, APCE 值将急剧减小。

当 APCE 与 F_{\max} 均低于一定阈值时,模型拒绝更新,既避免污染原模型,还降低了模型的更新次数,从而在一定程度上加快了跟踪算法的更新速度。当目

标连续多帧未更新时,考虑对跟踪算法重新初始化。

3.3 自适应模型更新

在相关滤波跟踪算法中,模型更新率的选择对算法跟踪性能的影响显著。模型的更新公式为

$$\begin{cases} a_i = (1 - \gamma)a_{i-1} + \gamma\hat{a} \\ \mathbf{K}_i = (1 - \gamma)\mathbf{K}_{i-1} + \gamma\mathbf{K}_{xx} \end{cases}, \quad (15)$$

式中: \mathbf{K}_i 为当前时刻核自相关估计矩阵; \mathbf{K}_{i-1} 为上一时刻核自相关估计矩阵; γ 控制模型的更新速率,当 γ 过大时,模型更新率加快,容易丢失目标原有特征而记忆背景干扰信息,导致错误跟踪;当 γ 过小时,难以实现对快速运动目标的跟踪。在标准目标跟踪数据集 OTB^[20] 视频序列中对跟踪算法进行分析,图 5 展示了不同模型更新率对更新效果的影响。

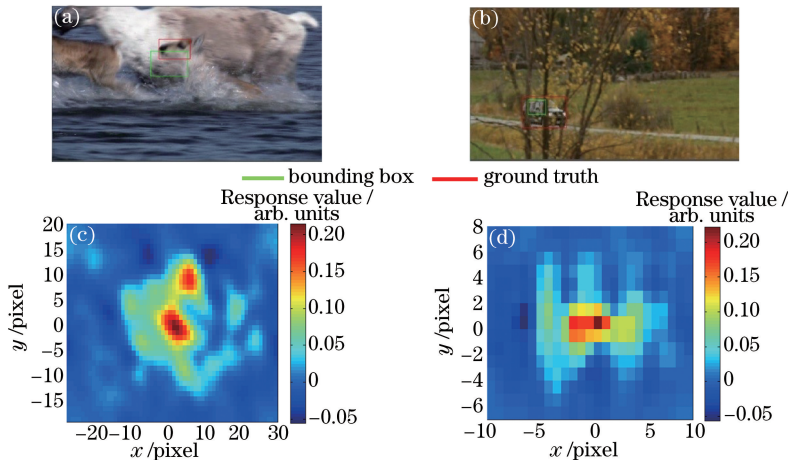


图 5 (a) 模型更新率偏小的序列; (b) 模型更新率偏大的序列; (c) 模型更新率偏小序列的响应图; (d) 模型更新率偏大序列的响应图
Fig. 5 (a) Sequence with small updating rate; (b) sequence with large updating rate; (c) response of sequence with small updating rate; (d) response of sequence with large updating rate

图 5(a)、(c)为模型更新率偏小序列及其对应的响应图,它们展示了被跟踪目标快速运动时模型更新率过小导致跟踪失败的情况。图 5(c)中存在两个峰值响应点,上面的峰值点为正确响应点,下面的峰值点为历史响应点。由于模型更新率过低,当目标发生快速运动时,算法因未及时更新模型而错误采纳了历史响应值,出现跟踪失败的问题。图 5(b)、(d)为模型更新率偏大序列及其对应的响应

图。示例中目标出现错误跟踪,这一方面源于目标出现了尺度变化,另一方面则由于距离较远时被跟踪目标在画面中相对运动缓慢,当出现障碍物遮挡时,过大的模型更新率及缓慢的移动容易将障碍物特征融入模型中,从而导致算法对障碍物的错误跟踪。

针对上述所存在的问题,引入 APCE 参数,对模型更新率进行动态调整:

$$\left\{ \begin{array}{l} \gamma = \gamma_{up}, \quad f_{up} < F_{max} \\ \gamma = \gamma_{down} + \frac{E_i}{E_h}(\gamma_{up} - \gamma_{down}), \quad f_{down} < F_{max} \leq f_{up} \quad \text{and} \quad E_i > \alpha E_h, \\ \gamma = 0, \quad F_{max} < f_{down} \quad \text{or} \quad E_i < \alpha E_h \end{array} \right. \quad (16)$$

式中: γ_{up} 为更新率上界值; γ_{down} 为更新率下界值; f_{up} 为响应上界值; f_{down} 为响应下界值; E_i 为当前响应图 APCE 值; E_h 为历史 APCE 值; α 为当前 APCE 值相对历史 APCE 值的下降比率。第 3.2 节已指出,APCE 指标反映了响应图起伏程度与当前

跟踪目标的置信程度。引入 APCE 参数在更新率上下界做线性插值,根据目标的最大响应值及响应图波动情况调节模型更新率。

本文算法框架示意图如图 5 所示。

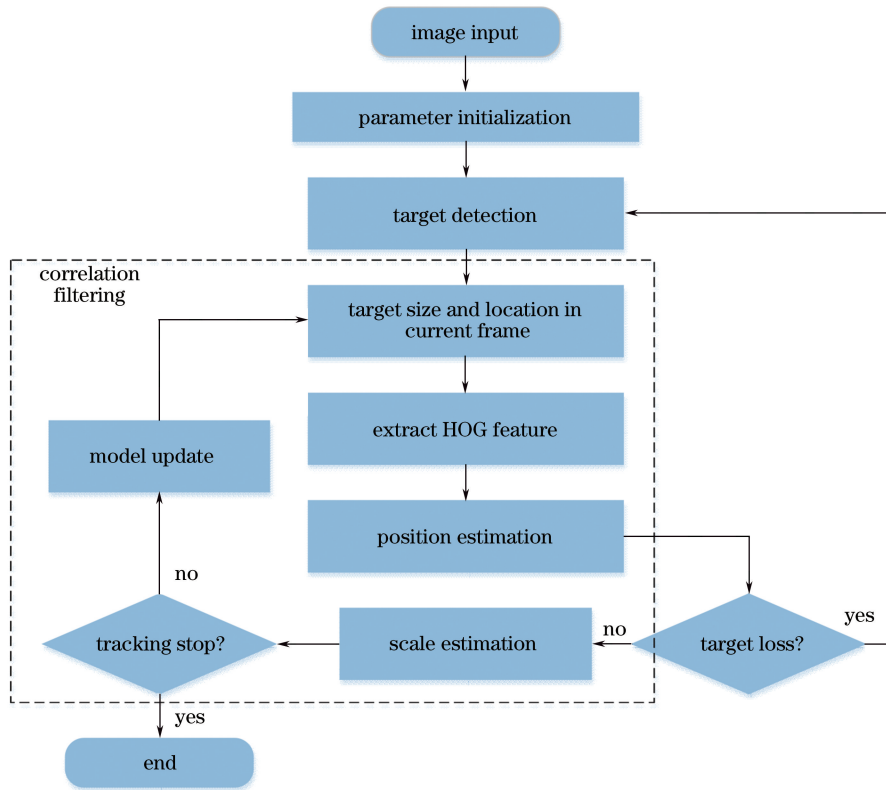


图 6 本文算法的流程示意图
Fig. 6 Flow chart of proposed algorithm

4 实验结果与分析

4.1 实验环境与评估指标

为了验证算法的可行性,采用 Matlab R2017b

开发平台,实验硬件配置为 core i7-6700 CPU 3.4 GHz, 16 GB 内存。相关滤波算法采用 OTB50^[20]数据集,包括光照变化、动态模糊、尺度变换等场景。

为了评估跟踪算法性能,根据 OTB 数据集建

议,选择了 OPE(one pass evaluation)模式,并采用了三种常见的评估指标:平均帧率(FPS)、距离精度(DP)和 OP。FPS 为每秒的运行帧数;DP 指标描述了预测框与真实框中心距离小于一定阈值(一般为 20 pixel)的序列占总序列的百分比;OP 指标描述了预测框与真实框交叠部分与二者面积的并集之比,即交叠比在 0.5 及以上的序列占总序列的百分比。

为检验本文算法的性能,将进行以下几个实验:1)目标丢失检测实验,针对目标跟踪状态与 APCE 指标间的关系进行分析;2)分析固定学习率与自适应学习率对目标跟踪的不同影响,并对其性能进行对比;3)将本文算法与其他常见的跟踪算法进行对比,通过上述给出的指标评价各个算法的优劣。

4.2 目标丢失检测实验

图 7 给出了遮挡和变形场景下最大响应值和 APCE 指标的变化情况。图 7(a)、(b)为 woman2

序列及其响应图,图 7(c)、(d)为 tiger2 序列及其响应图。两个序列均出现了目标遮挡情况,其中图 7(b)在目标遮挡的情况下,依旧有较大的峰值响应, $F_{\max}=0.27$ 。图 7(e)、(f)为 Sylvester 序列及其响应图,该序列出现了目标形变。从图 7(b)可知,单纯响应最大值并不能反映响应图的波动情况。采用 APCE 指标可有效反映响应图的振荡情况,当目标被遮挡、丢失或者发生形变时,APCE 相对序列 APCE 历史均值显著减小,在此情形下拒绝模型更新,避免模型被污染。

实验过程中,高峰值响应依旧可以作为目标丢失检测的重要判据,尤其当 $F_{\max}>0.5$ 时,可认为目标被成功跟踪。这里采用的更新策略为:当 $F_{\max}>0.5$ 时,认定目标已成功跟踪;当 $0.5 \geq F_{\max} > 0.2$ 时,采用 APCE 对响应图振荡程度作出判断;当 F_{\max} 低于一定历史均值时,拒绝模型更新;将图像的响应峰值与 APCE 指标相结合,一方面可减少运算量,另一方面可提高判断跟踪成功的准确度。

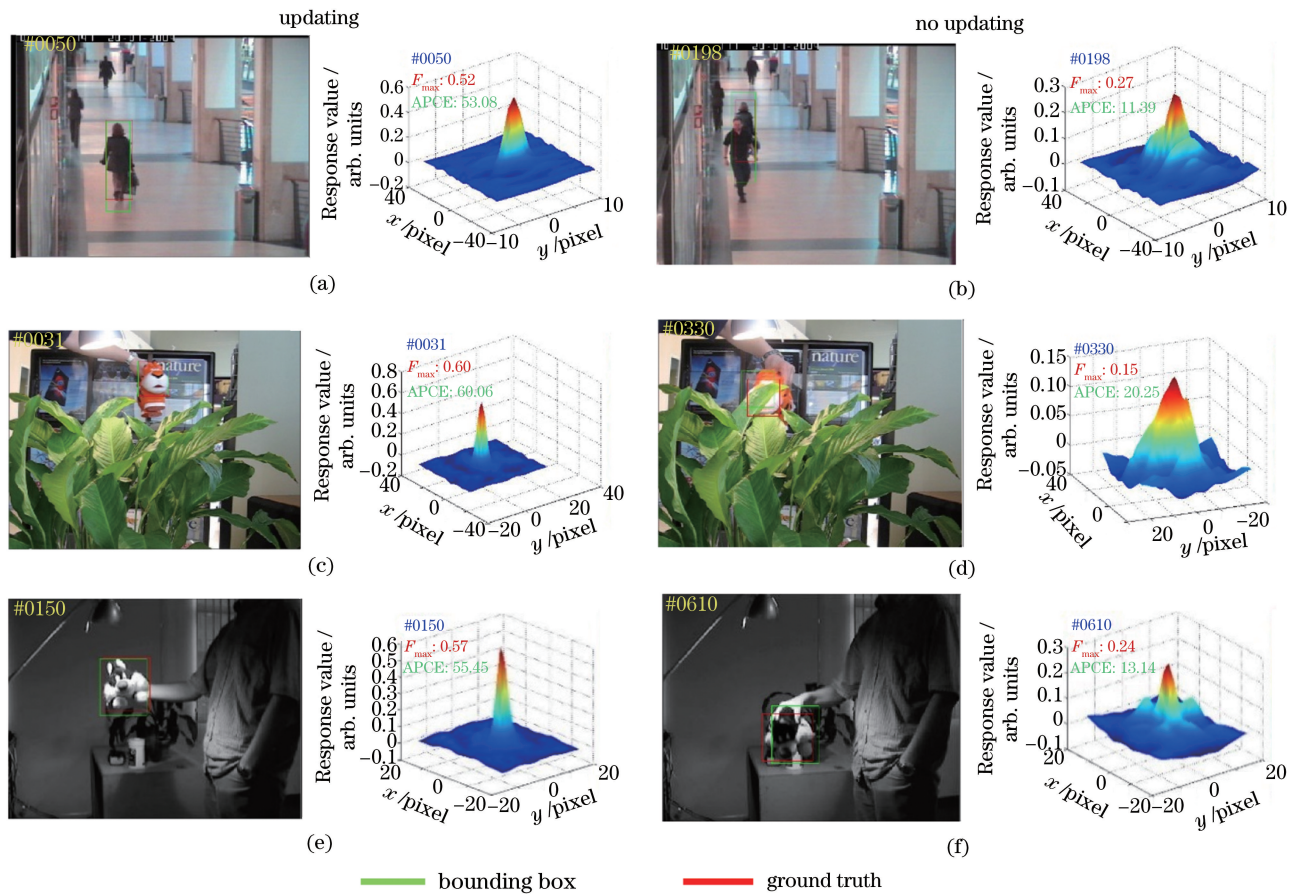


图 7 遮挡和变形场景下最大响应值和 APCE 指标的变化情况。(a)(b) woman2 序列及其响应图;(c)(d) tiger2 序列及其响应图;(e)(f) Sylvester 序列及其响应图

Fig. 7 Variation of maximum response value and APCE index in occlusion or deformation scene. (a)(b) woman2 sequence and its response map; (c)(d) tiger2 sequence and its response map; (e)(f) Sylvester sequence and its response map

4.3 自适应模型更新对比

参考 KCF^[7]、DSST^[10] 等算法的取值, (16) 式中, $\gamma_{up}=0.03, \gamma_{down}=0.01, f_{up}=0.5, f_{down}=0.2, \alpha=0.6$ 。图 8 比较了对 tiger2 序列的固定学习率与自适应更新率。初始阶段, 采用固定更新率的跟踪框与采用自适应更新率的跟踪框都保持较高的跟踪精度。图 8(a) 目标未出现遮挡, 图 8(b)~(d) 中虽目标出现遮挡, 但依旧可识别出所要跟踪的对象。在

前期跟踪过程中, 固定学习率算法不加区分地学习引入障碍物特征, 从而错误地识别跟踪对象 [图 8(e)], 并彻底丢失跟踪目标特征 [图 8(f)]。采用固定更新率的算法, DP(20 pixel 误差内) 精度为 0.85, OP 精度指标为 0.86 (检测评价函数 IOU 为 0.5)。相比而言, 采用自适应更新率算法拥有更高跟踪精度, DP(20 pixel 误差内) 精度为 0.97, OP 精度为 0.98 (检测评价函数 IOU 为 0.5)。

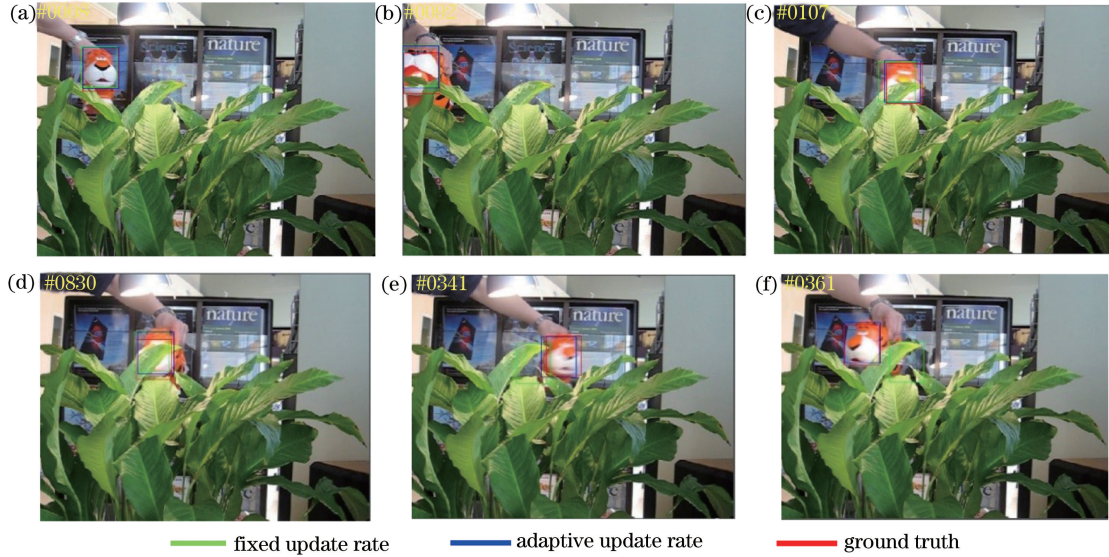


图 8 tiger2 序列的固定更新率与自适应更新率的比较。(a) 图像未遮挡; (b)~(d) 图像被遮挡; (e) 错误识别图像; (f) 丢失跟踪目标特征

Fig. 8 Comparison between fixed updating rate and adaptive updating rate for tiger2 sequence. (a) Unoccluded image; (b)-(d) occluded image; (e) misidentified image; (f) lost tracking target characteristics

4.4 视觉跟踪算法对比

为了验证所提出的算法, 同时对比和评估不同跟踪算法的性能和效果, 选取了其他 10 种流行的算法: DSST^[10]、KCF^[7]、Staple^[21]、SAMF^[22]、CSK^[6]、TLD^[5]、Struck^[23]、MTT^[24]、LOT^[25] 和 CPF^[26]。

图 9 给出了 11 种算法在 OTB 数据集 50 段图像序列中的 OP 和 DP 曲线。由图可知, 由于本文

算法、Staple、SAMF 和 DSST 算法引入了自适应尺度估计, 在跟踪精度上优于其他非尺度估计算法; 在包含尺度估计的算法中, 本文算法采用了自适应更新率和丢失目标重检测机制, 跟踪精度略高于其他算法; 本文算法的平均 OP 和平均 DP 与 SAMF 和 Staple 算法基本持平; 在 FPS 方面, 本文算法采用相关滤波框架体系, 其运算速度远高于其他非相关

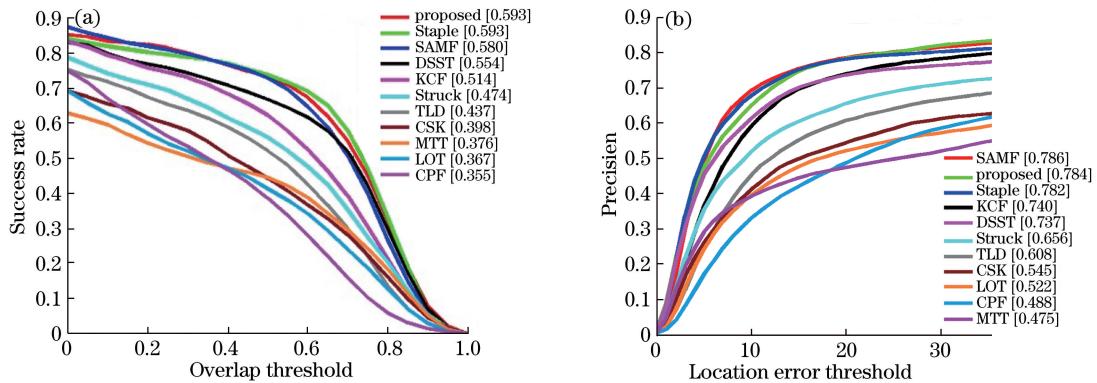


图 9 不同算法的 (a) OP 曲线和 (b) DP 曲线

Fig. 9 (a) OP curves and (b) DP curves by different algorithms

滤波算法,并且高于采用相关滤波的跟踪算法(Staple, DSST, KCF 和 SAMF),其中 KCF 算法拥有 354.9 frame/s 的超高速率,但是没有引入尺度估计,精度较低;SAMF 算法在 OP 和 DP 方面表现出色,但是较高的算法复杂度导致其 FPS 低。

图 10 是 11 种跟踪算法在部分序列上的实际跟踪效果图。其中,在 singer2 序列发生尺度变换、光照变化、连续形变的情况下,本文算法、Staple、DSST 和 KCF 算法始终跟随目标,验证了本文算法

在尺度变化场景上的有效性;skating1 序列在前述的干扰因素中,额外引入了垂直轴向旋转,在此情景下,本文算法、Staple、SAMF、DSST、KCF 算法保持了较好的跟踪精度;soccer 序列则存在运动模糊、背景混杂和垂直轴向旋转等因素,在此场景下,本文算法、DSST、KCF、LOT 算法表现出了较好的跟踪性能,这反映了本文算法具有很好的稳健性。本文算法、Staple、SAMF、DSST、KCF 和 TLD 算法的 FPS 依次为 65.3, 67.8, 25.1, 60.5, 354.9, 22.4 frame/s。

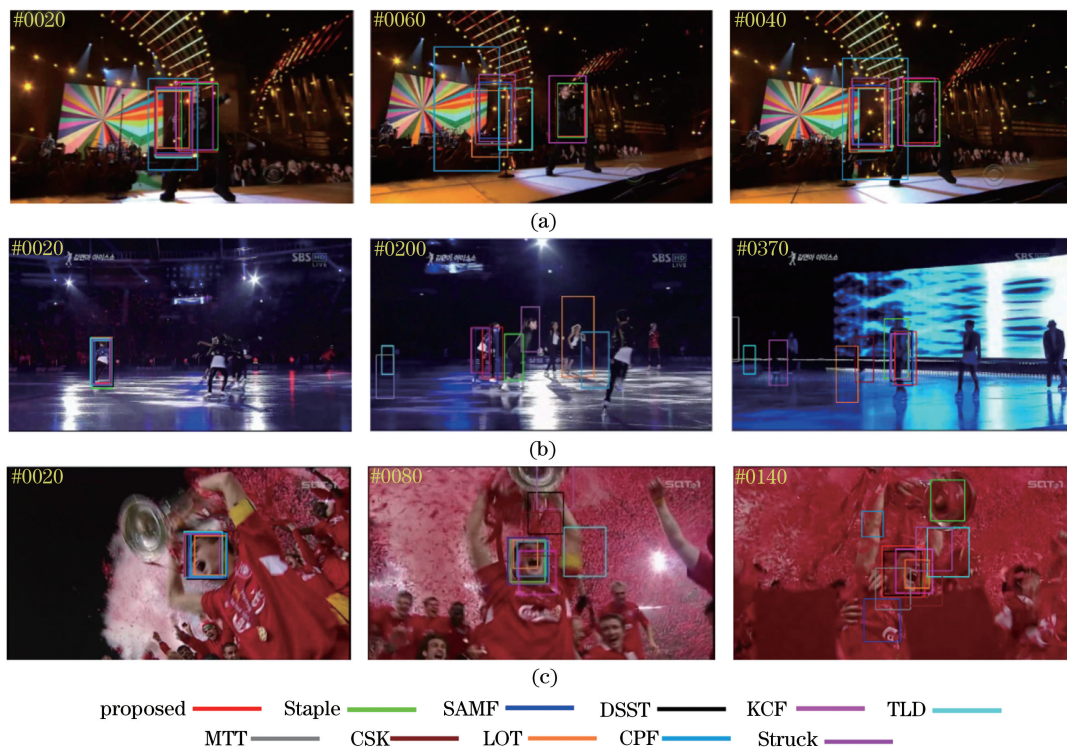


图 10 本文算法与其他 10 种算法的部分跟踪结果对比。(a) singer2 序列;(b) skating1 序列;(c) soccer 序列
Fig. 10 Comparison of partial tracking results by proposed algorithm and other ten algorithms. (a) singer2 sequence;
(b) skating1 sequence; (c) soccer sequence

5 结 论

实现了一种结合目标检测的相关滤波视觉跟踪算法。首先利用基于深度学习的目标检测算法得到目标的位置、大小信息,并利用相关滤波算法对所给出的目标进行视觉跟踪。当目标响应值异常时,停止对模型进行更新,当连续数帧目标响应值异常,对所跟踪目标重新初始化,形成检测-跟踪-检测的闭环视觉跟踪过程。本文算法既解决了传统跟踪算法需要提供初始跟踪区域的难题,又解决了相关滤波跟踪误差随时间积累的问题。相比于基于实时检测算法实现的目标跟踪,本文算法的运行速度更快,且具有较好的稳健性。在 Matlab 平台下,得到本文算法的平均定位精度为 0.593,平均 OP 为 0.784,FPS

为 65.3。

参 考 文 献

- [1] Zhang X G, Zhang Y, Wang Y N, *et al.* Covariance tracking based on forgetting factor and Kalman filter [J]. Acta Optica Sinica, 2010, 30(8): 2317-2323.
张旭光, 张云, 王艳宁, 等. 基于遗忘因子与卡尔曼滤波的协方差跟踪[J]. 光学学报, 2010, 30(8): 2317-2323.
- [2] Liu W, Zhao W J, Li C. Long-term visual tracking based on spatio-temporal context [J]. Acta Optica Sinica, 2016, 36(1): 0115001.
刘威, 赵文杰, 李成. 时空上下文学习长时目标跟踪 [J]. 光学学报, 2016, 36(1): 0115001.
- [3] Cheng Y Z. Mean shift, mode seeking, and

- clustering[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790-799.
- [4] Bradski G R. Computer vision face tracking for use in a perceptual user interface[C]. IEEE Workshop on Applications of Computer Vision, 1998: 214-219.
- [5] Kalal Z, Mikolajczyk K, Matas J. Face-TLD: Tracking-learning-detection applied to face[J]. Proceedings of the IEEE, 2010, 119(5): 3789-3792.
- [6] Henriques J F, Caseiro R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels[C]. European Conference on Computer Vision, 2012: 702-715.
- [7] 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.
- [8] Bolme D S, Beveridge J R, Draper B A, *et al.* Visual object tracking using adaptive correlation filters[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 2544-2550.
- [9] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005: 886-893.
- [10] 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.
- [11] Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4800-4808.
- [12] 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.
- [13] Xu Y L, Wang J B, Li Y, *et al.* Scale-adaptive tracking based on kernelized correlation filter[J]. Application Research of Computers, 2016, 33(11): 3513-3516, 3520.
徐玉龙, 王家宝, 李阳, 等. 基于相关滤波的尺度自适应目标跟踪[J]. 计算机应用研究, 2016, 33(11): 3513-3516, 3520.
- [14] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [15] Girshick R. Fast R-CNN[J]. IEEE International Conference on Computer Vision and Pattern Recognition, 2015: 1440-1448.
- [16] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [17] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [18] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector[C]. European Conference on Computer Vision, 2016: 21-37.
- [19] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. Computer Science, 2014: arXiv.
- [20] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.
- [21] Bertinetto L, Valmadre J, Golodetz S, *et al.* Staple: complementary learners for real-time tracking [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1401-1409.
- [22] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracked with feature integration[C]. European Conference on Computer Vision, 2014: 254-265.
- [23] Hare S, Saffari A, Torr P. Struck: structured output tracking with kernels[C]. IEEE International Conference on Computer Vision, 2011: 263-270.
- [24] Zhang T Z, Ghanem B, Liu S, *et al.* Robust visual tracking via multi-task sparse learning [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012: 2042-2049.
- [25] Oron S, Bar-Hillel A, Levi D, *et al.* Locally orderless tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2012: 1940-1947.
- [26] Perez P, Hue C, Vermaak J, *et al.* Color-based probabilistic tracking [C]. European Conference on Computer Vision, 2002: 661-675.