

基于相关滤波融合多特征的运动目标跟踪方法

谢柳 尚振宏 刘辉

(昆明理工大学信息工程与自动化学院, 昆明, 650500)

摘要: 针对复杂环境下仅使用单一图像特征跟踪精度和鲁棒性差的问题, 提出一种多特征融合的相关滤波目标跟踪算法。该算法首先从目标和背景区域分别提取方向梯度直方图(Histogram of oriented gradient, HOG)特征、颜色直方图特征和卷积特征, 采用固定权重方法融合HOG特征和颜色直方图特征的特征响应图, 然后将该层融合结果与卷积特征响应图采用自适应权重融合策略进行融合, 基于融合后的响应图估计出目标位置, 并采用尺度估计方法解决目标尺度变化问题, 最后采用稀疏模型更新策略进行模型更新。在OTB-2013公开标准测试集中验证本文算法性能, 并与主流的目标跟踪算法进行了对比分析。实验结果表明, 与其中最优算法相比, 本文算法的平均距离精度值和平均重叠精度值都有所提高。本文算法由于有效地利用了HOG特征、颜色直方图特征和卷积特征, 在复杂场景下目标跟踪的准确性和鲁棒性都优于其他算法。

关键词: 目标跟踪; 相关滤波; 多特征融合; 自适应权重

中图分类号: TP391.41 **文献标志码:** A

Object Tracking Combining Multiple Features Based on Correlation Filter

Xie Liu, Shang Zhenhong, Liu Hui

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, 650500, China)

Abstract: Aiming at the problem that object tracking with single image feature under complex circumstances has low accuracy and poor robustness, a correlation filtering object tracking algorithm based on multi-feature fusion is proposed. Firstly, histogram of oriented gradient (HOG) features, color histogram features and convolutional features are respectively extracted from the target and background regions, and a fixed-coefficient fusion strategy is adopted to combine the feature response maps of HOG features and color histogram features. Then the fused response map and the convolutional features response map are fused by adaptive weighted fusion strategy, and the scale estimation algorithm is used to solve the problem of target scale changes. Finally, the sparse model update strategy is used to update the model. The proposed algorithm is evaluated on OTB-2013 dataset and compared with state-of-the-arts object tracking algorithms. Extensive experimental results show that our method significantly improves the performance in median distance precision and median overlap precision compared to the optimal algorithm. The accuracy and robustness of the proposed algorithm are superior to those of other algorithms in complex scenarios because of the effective use of HOG, color histogram and convolutional features.

Key words: object tracking; correlation filter; multi-feature fusion; adaptive weight

引言

在计算机视觉领域中,目标跟踪一直是一个重要课题^[1],在军事和民用方面都有广阔的应用前景,主要包括人机交互、军事制导、视频监控和智能交通等。虽然近年来目标跟踪问题已经取得了很大的突破,但由于跟踪环境的复杂多样性,如尺度变化、光照变化、目标旋转、遮挡与背景干扰等^[2],实现鲁棒的目标跟踪仍是一个极具挑战的难题。

目标跟踪方法按照目标外观模型建立的方式不同可分为生成式模型和判别式模型。生成式方法是对目标建立目标外观模型,然后在后续帧中搜索目标候选区域并与目标外观模型进行对比,得到与之相似度最高的跟踪结果。由于生成式方法仅仅对目标本身进行描述,没有考虑背景信息,因此跟踪效果并不理想。判别式方法是对目标区域和背景区域分别提取样本来训练目标分类器,然后在后续帧利用训练好的分类器来预测目标的位置,进而完成目标跟踪。

近几年,基于相关滤波的判别式目标跟踪算法由于具有较好的实时性和鲁棒性,引起了相关学者的广泛关注。Bolme等^[3]首次将相关滤波思想应用到目标跟踪领域,提出了最小化输出平方误差和(Minimum output sum of square error, MOSSE)的目标跟踪算法。Henriques等^[4]在相关滤波基础上引入核函数,提出核相关滤波器(Kernel correlation filter, KCF)的目标跟踪算法,并结合循环矩阵理论和快速傅里叶变换,解决了训练样本不足和计算量大的问题。随后很多研究工作在其基础上进行改进,提出了多种改进的相关滤波跟踪算法。

目标跟踪中目标的特征提取和表示是实现目标鲁棒跟踪的关键因素之一。早期 MOSSE 和 KCF 算法分别提取目标的灰度特征和方向梯度直方图(Histogram of oriented gradient, HOG)特征进行目标跟踪,取得了实时跟踪效果,但跟踪精度有待提高。随后, Danelljan 等分别使用融合 HOG 特征和灰度特征表征目标的方法^[5],以及颜色属性^[6](Color name, CN)表征目标的方法进行目标跟踪,其算法跟踪精度都有所提升。Li 等^[7]提出融合 HOG 和 CN 特征的目标跟踪算法,很大程度上提高了跟踪精度。Bertinetto 等^[8]提取目标的 HOG 特征和颜色直方图特征对目标进行描述,实现了目标的鲁棒跟踪。随着深度学习的快速发展,卷积神经网络(Convolutional neural network, CNN)在目标检测和识别等领域取得了巨大成功,受此启发,众多研究者开始将深度学习网络框架应用到目标跟踪领域。Wang 等^[9]首次利用深度学习框架,通过堆栈降噪自编码器(Stacked denoising auto-encoder, SDAE)训练获得通用目标特征提取框架的方式进行目标跟踪,取得了比传统跟踪算法更好的效果。Ma 等^[10]提出基于分层卷积特征的跟踪算法(Hierarchical convolutional features for visual tracking, HCF/CF2),利用预训练好的 CNN 模型提取不同深度的卷积特征作为目标特征表示,跟踪精度得到了很大的提升。Qi 等^[11]提出多层卷积相关滤波跟踪算法(Hedged deep tracking, HDT),通过自适应学习权重策略融合多层卷积特征进行目标跟踪,使跟踪精度提高,但跟踪速度很慢。Wang 等^[12]提出全卷积目标跟踪算法(Fully convolutional networks tracking, FCNT),通过将高层卷积特征和底层卷积特征分别送入一般性网络(General network, GNet)和特殊性网络(Specific network, SNet)用来判断类间和类内目标类别,并生成各自的热度图谱,最后结合热度图谱确定目标位置,其跟踪精度高,但由于模型复杂,跟踪速度比较慢。Bertinetto 等^[13]提出一种全卷积孪生网络结构(Fully-convolutional siamese networks, SiamFC),通过将该结构提取的目标模板与候选区域特征进行卷积操作,得到目标响应图,响应最大候选区域位置即为目标位置,算法跟踪精度和速度都有所提升。

除了各种特征的组合使用和卷积特征的改进外,也有不少算法通过对尺度估计、滤波器本身以及模型更新等其他辅助方法进行改进,从而提升跟踪算法的鲁棒性。针对目标尺度发生变化的问题, Danelljan 等^[5]提出一种判别式尺度空间跟踪算法(Discriminative scale space tracker, DSST),训练尺度相关滤波器来估计目标尺度,算法简单高效。Li 等^[7]采用尺度自适应的多特征融合跟踪算法(Scale adap-

tive with multiple features tracker, SAMF), 采用构造目标图像尺度池的方法搜索目标最佳尺度。为了解决相关滤波器本身存在的问题, Danelljan 等^[14]提出空间正则化的相关滤波方法(Spatially regularized discriminative correlation filter, SRDCF), 在损失函数中引入空间正则化惩罚项, 从而减小循环平移引起的边缘效应。Danelljan 等^[15]提出连续空间域的卷积跟踪算法(Continuous convolution operators for visual tracking, CCOT), 目标定位时可达到亚像素精度的位置, 跟踪精度高, 但跟踪速度极慢。针对跟踪速度慢, 模型在遮挡情况易受污染等问题, Zhao 等^[16]引入峰值旁瓣比(Peak-to-sidelobe ratio, PSR)作为相关滤波响应检测指标, 提出一种稀疏更新模型策略, 从而解决遮挡情况下的模板更新问题, 进一步提高了目标跟踪的速度和精度。Wang 等^[17]通过平均峰值相关能量(Average peak-to-correlation energy, APCE)判断是否出现目标遮挡或丢失, 根据其值自适应对目标模型进行更新, 改善了算法的鲁棒性和实时性。

上述基于目标特征选取改进的跟踪算法仅使用单一的融合策略对传统特征进行融合, 由于在复杂环境下使用单一的特征融合策略鲁棒性不强, 跟踪算法易受跟踪环境变化的干扰, 无法准确跟踪目标。另外, 手工设计的传统特征具有一定的局限性, 不能有效捕捉目标物体的语义信息, 难以处理复杂的目标表现变化。为进一步提高目标跟踪的鲁棒性, 本文提出了一种分层融合 HOG 特征、颜色直方图特征和卷积特征的相关滤波目标跟踪算法。另外, 受上述其他改进算法的启发, 为解决目标尺度变化问题以及多种特征的组合使用导致速度缓慢的问题, 本文同样采用尺度估计方法和稀疏更新模型策略来提高算法的稳定性和实时性。

1 相关滤波跟踪算法

本文在相关滤波基础上融合了 HOG 特征、颜色直方图特征及卷积特征, 为体现论文完整性, 将对相关滤波跟踪算法简述如下, 其详细论述见参考文献[5]。

记 d 通道目标外观模板为 x , 其第 d 个通道特征表示为 $x^d, d \in \{1, 2, \dots, D\}$ 。记相关滤波器为 w , 由 D 个单通道滤波器 w^d 组成。相关滤波器可通过最小化目标函数 ϵ 求解, 即

$$\epsilon = \left\| \sum_{d=1}^D w^d * x^d - y \right\|^2 + \lambda \sum_{d=1}^D \|w^d\|^2 \quad (1)$$

式中: $*$ 代表循环相关操作, y 表示相关滤波期望输出, λ 为正则化参数。利用傅里叶变换快速求解可得到相关滤波器^[18]为

$$W^d = \frac{\bar{Y} X^d}{\sum_{d=1}^D \bar{X}^d X^d + \lambda} \quad (2)$$

式中: 大写字母表示相应变量的频域描述形式, \bar{Y} 为 Y 的复共轭, $\bar{Y}X$ 表示对应元素相乘。对下一帧新的图像块, 提取图像块特征 z^d , 计算傅里叶变换后的特征 Z^d 与滤波器 W^d 的相关输出响应 f 为

$$f = F^{-1} \left(\sum_{d=1}^D \bar{W}^d Z^d \right) \quad (3)$$

式中: F^{-1} 表示傅里叶逆变换。输出响应 f 取得的最大值处即为新的目标位置。

在目标跟踪过程中, 目标的外观会发生变化, 为了能持续跟踪目标, 滤波器需要在线更新, 具体更新方式为将式(2)拆分成分子 A_t^d 和分母 B_t 分别进行, 即

$$\begin{aligned} A_t^d &= (1 - \eta) A_{t-1}^d + \eta_{\text{corr}} \bar{Y} X_t^d \\ B_t &= (1 - \eta) B_{t-1} + \eta_{\text{corr}} \sum_{d=1}^D \bar{X}_t^d X_t^d \end{aligned} \quad (4)$$

式中: t 为当前帧的序列号, η_{corr} 为学习率。

2 本文算法

本文算法在相关滤波框架下,采用HOG特征、颜色直方图特征和卷积特征进行目标跟踪。首先采用固定系数融合策略进行HOG和颜色直方图的特征响应图融合,将该层融合结果与基于卷积特征获得的特征响应图进行自适应加权融合,然后利用融合后的响应图进行目标位置估计,并基于估计的目标位置采用尺度估计方法进行目标尺度估计,最后采用稀疏更新策略进行模型更新。

2.1 多特征训练滤波器

2.1.1 HOG特征和颜色直方图特征

HOG特征是图像的局部特征,通过计算图像局部区域的方向梯度直方图来构成特征,对目标细小形变、光照变化等有较好的适应能力,但是当目标发生较大形变或遮挡时,HOG特征不能适应目标的变化从而丢失目标。与HOG特征不同的是,颜色特征是一种基于像素点的全局特征,描述图像的全局外观特征,能够有效处理目标形变和尺度变化问题,但是颜色特征无法适应光照变化。为此,本文将这两种特征融合起来描述目标模型,可以很好地弥补各自的缺点,提高目标跟踪算法的准确性和鲁棒性。对于采用颜色直方图特征作为目标特征表示的方法,Staple算法^[8]首次将其引入到了相关滤波跟踪算法中,具体方法简述如下。

对于给定视频帧,先获取图像块的目标区域 O 和背景区域 B ,在目标和背景区域的每个像素上,通过岭回归模型进行训练,其训练目标函数 L_{color} 为

$$L_{\text{color}} = \frac{1}{|O|} \sum_{u \in O} (\beta^T \varphi[u] - 1)^2 + \frac{1}{|B|} \sum_{u \in B} (\beta^T \varphi[u])^2 \quad (5)$$

式中:颜色直方图滤波器为 β , $|O|$ 和 $|B|$ 分别表示目标区域内和背景区域内的像素数目; $\varphi[u] = e_{k[u]}$,表示目标特征,其中 $e_{k[u]}$ 是长度为 V 的向量,其元素只在下标为 $k[u]$ 的位置为1,其他位置为0; $k[u]$ 表示位置 u 处像素对应的直方图区间值; V 是颜色直方图的区间数目。

利用 $\beta^T \varphi[u] = \beta^T e_{k[u]} = \beta^{k[u]}$,通过对 β 的每一特征维度进行求解,得到 β 为

$$\beta^c = \frac{\rho^c(O)}{\rho^c(O) + \rho^c(B) + \tau} \quad (6)$$

式中: $c = 1, \dots, V$; $\rho(O)$ 和 $\rho(B)$ 分别为目标和背景的颜色直方图,且 $\rho^c(A) = \frac{N^c(A)}{|A|}$, $N^c(A)$ 表示区域 A 中属于颜色直方图区间 c 的统计数目。利用式(6)求解 β ,通过积分图技术^[19]计算得到颜色直方图特征响应图 f_{color} 。

和相关滤波器一样,为了适应目标外观变化而能持续跟踪目标,采用滑动平均方法分别对 $\rho(O)$ 和 $\rho(B)$ 进行在线更新,具体更新方式为

$$\begin{aligned} \rho_t(O) &= (1 - \eta_{\text{color}}) \rho_{t-1}(O) + \eta_{\text{color}} \rho_t(O) \\ \rho_t(B) &= (1 - \eta_{\text{color}}) \rho_{t-1}(B) + \eta_{\text{color}} \rho_t(B) \end{aligned} \quad (7)$$

式中: t 为当前帧的序列号, η_{color} 为学习率。

2.1.2 卷积特征

传统相关滤波跟踪算法^[3-8]都是利用人工设计特征训练相关滤波器,跟踪精度有待进一步提高。由于卷积特征表征能力强,本文利用卷积特征作为输入特征,采用相关滤波的方法对目标外观模型进行训练获得相关滤波器。

根据文献^[10]所述,在VGG-19网络中,低层的卷积特征能够更好地保留目标自身的位置和空间信息,深层的卷积特征却包含更多的语义信息。图1分别给出了输入图像在conv3-4, conv4-4和conv5-4

中的卷积特征。如图1所示,conv3-4输出的卷积特征保留更多的空间位置信息,conv4-4输出的卷积特征包含部分的空间信息和语义信息,而conv5-4输出的卷积特征包含更多的语义信息,因此分别对conv3-4,conv4-4和conv5-4输出的卷积特征训练相关滤波器。具体过程描述如下。

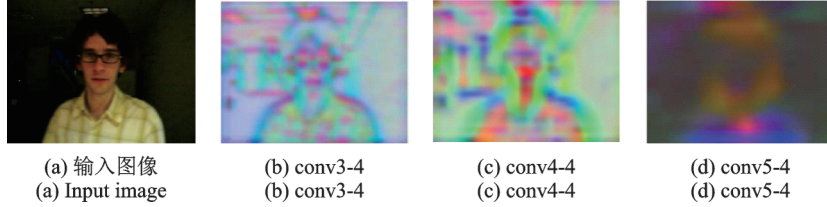


图1 目标在不同卷积层中的特征表示

Fig.1 Feature representations of object at different convolutional layers

记目标样本为 x , 提取第 l 层的第 d 个通道的卷积特征 $x_l^d, l \in \{1, 2, 3\}$, 其中 1 表示 conv3-4 层, 2 表示 conv4-4 层, 3 表示 conv5-4 层, $d \in \{1, 2, \dots, D\}$ 。记第 l 层相关滤波器为 w_l , 由 D 个单通道滤波器 w_l^d 组成。相关滤波器可通过最小化式(1)求解, 利用傅里叶变换快速求解可得到相关滤波器 W_l^d , 如式(2)所示。对下一帧新的图像块, 提取图像块的卷积特征 z_l^d , 通过式(3)计算傅里叶变换后的特征 Z_l^d 与滤波器 W_l^d 的相关输出响应 f_l 。

高层卷积特征语义信息丰富, 对目标外观特征表示较好, 应该给予较高的融合权重; 低层卷积特征分辨率较高, 能够精确定位目标, 但是对遮挡、形变比较敏感, 应该给予较低的融合权重。按照上述原则, 对每层的响应图加权相加得到最终卷积特征响应图 f_{conv} , 即

$$f_{\text{conv}} = \sum_{l=1}^3 \alpha_l f_l \quad (8)$$

式中 α_l 表示每层对应的融合权重。

2.2 多特征自适应融合

根据上述不同类型特征描述目标的能力不同, 应对不同特征采取分层融合策略。首先, 采用固定权重融合策略进行 HOG 特征和颜色直方图特征的响应图融合, 融合公式为

$$f_{\text{trad}} = (1-r)f_{\text{hog}} + rf_{\text{color}} \quad (9)$$

式中: r 为固定权重, 其值设置参考文献[8]。其次, 由于 HOG 特征和颜色直方图特征都属于传统人工设计特征, 相比卷积特征, 对不同场景下目标的表征不够丰富。为了充分利用传统特征和卷积特征, 文中将两种传统特征融合得到的响应图 f_{trad} 和卷积特征响应图 f_{conv} 进行自适应加权融合得到最终目标响应图。

为计算传统特征和卷积特征在跟踪过程中的判别能力, 本文采用相邻图像帧的特征响应图 PSR 差值来计算特征融合权重。差异越小说明该特征下学习到的滤波器性能越好, 则在特征响应图融合中应对该特征响应图赋予更高的权重。因此, 两个特征各自的自适应权重分别为

$$\tilde{k}_t^{\text{conv}} = \frac{C_t^{\text{trad}}}{C_t^{\text{conv}} + C_t^{\text{trad}}}, \tilde{k}_t^{\text{trad}} = 1 - \tilde{k}_t^{\text{conv}} \quad (10)$$

式中: t 为当前帧的序列号, C_t^{trad} 和 C_t^{conv} 分别表示相邻帧的传统特征和卷积特征响应图峰值旁瓣比差值, 其计算公式为

$$\text{PSR}(f) = \frac{\max(f) - \mu}{\sigma}, C_t = \text{PSR}(f_t) - \text{PSR}(f_{t-1}) \quad (11)$$

式中: μ 和 σ 分别为响应图 f 的均值和方差。各自权重的更新策略为

$$k_t = (1 - \eta_k)k_{t-1} + \eta_k \tilde{k}_t \quad (12)$$

式中 η_k 为权重更新系数。通过计算分别得到传统特征和卷积特征的融合权重,然后对两特征响应图进行加权融合,得到最终的融合响应图 f_{final} 。

图2为本文多特征自适应融合方法在 Bolt 视频序列上运行的特征权重变化图。从图中可以看出,卷积特征权重在视频序列的大部分帧中都很大,尤其是背景杂乱环境对目标产生严重干扰时,譬如 Bolt 视频序列的第 137 帧,其他运动员衣服和广告牌的颜色干扰使得传统特征权重下降。但是,当目标快速运动致边界轮廓模糊或遇到相似目标时,卷积特征权重稍微有点下降,譬如 Bolt 视频序列的第 65 帧和第 107 帧。

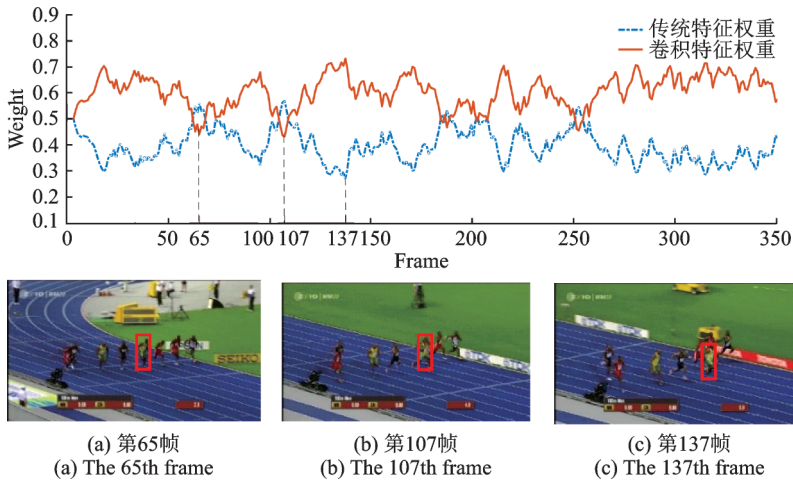


图2 Bolt 视频序列不同特征的融合权重变化曲线

Fig.2 Variation curve of fusion weights of different features on Bolt

2.3 尺度估计

针对目标尺度变化问题,本文采用文献[5]中尺度估计算法估计目标尺度,即:基于预测得到的目标位置建立目标尺度变化的“金字塔”。用 $M \times N$ 表示目标的尺度大小, S 表示尺度样本个数, a 表示尺度因子,以 $a^n M \times a^n N$ 为尺度来提取目标样本,其中 $n \in \left\{ \left[-\frac{S-1}{2} \right], \dots, \left[\frac{S-1}{2} \right] \right\}$,然后将得到的不同尺度的目标样本统一缩放为 $M \times N$ 的大小,并与一维尺度相关滤波器进行相关操作获得尺度响应图,其最大值位置所表示的尺度就是目标的最佳尺度。

2.4 模型更新

由于多种特征组合的使用使得跟踪的速度很慢,很难达到实时跟踪效果。考虑到在一段视频序列中,相邻两帧图片之间目标的变化很小,如果每一帧都对模型进行更新,容易造成时间浪费,并且当目标发生遮挡时,容易将错误的目标特征信息加到目标模型中而导致跟踪失败,因此本文引入 APCE 值^[17]来判断目标是否发生丢失或飘移,从而实现稀疏更新模型。图3为 Jogging1 视频序列的 APCE 值响应曲线图。由图可知,目标在第 7,240 帧时正常运动,APCE 值比较大;在第 79 帧时,目标被完全遮挡,APCE 值较小。

APCE 值用来表征目标响应图的波动程度和目标的置信程度,其计算公式为

$$APCE = \frac{|f_{\max} - f_{\min}|^2}{\text{mean}(\sum_{i,j} (f_{i,j} - f_{\min})^2)} \quad (13)$$

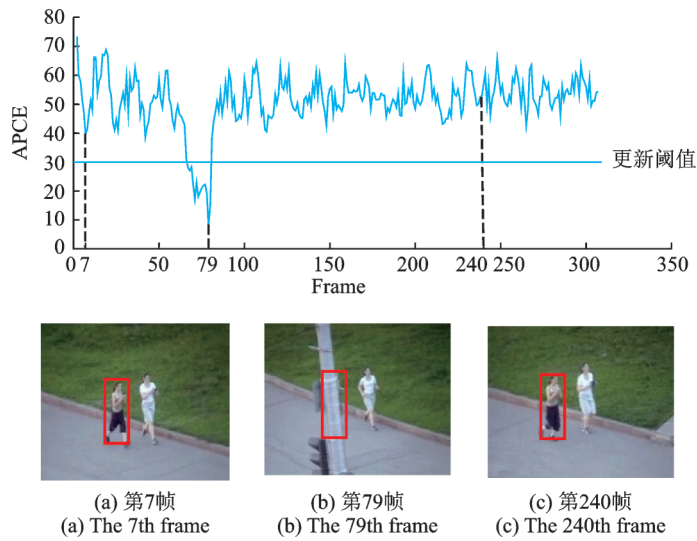


图3 Jogging1视频序列的APCE值变化曲线

Fig.3 Variation curve of APCE value on Jogging1

式中： f_{\max} 和 f_{\min} 分别表示响应图中的最大值和最小值， $f_{i,j}$ 表示响应图中第*i*行第*j*列的值。当响应图的APCE值大于更新阈值时，则该跟踪结果被认为是高置信度，那么跟踪模型分别通过式(4)和式(7)进行在线更新。

2.5 算法具体实施过程

本文算法的基本流程示意图如图4所示，主要分为4个过程：滤波器训练、位置估计、尺度估计和模型更新。具体步骤如下。

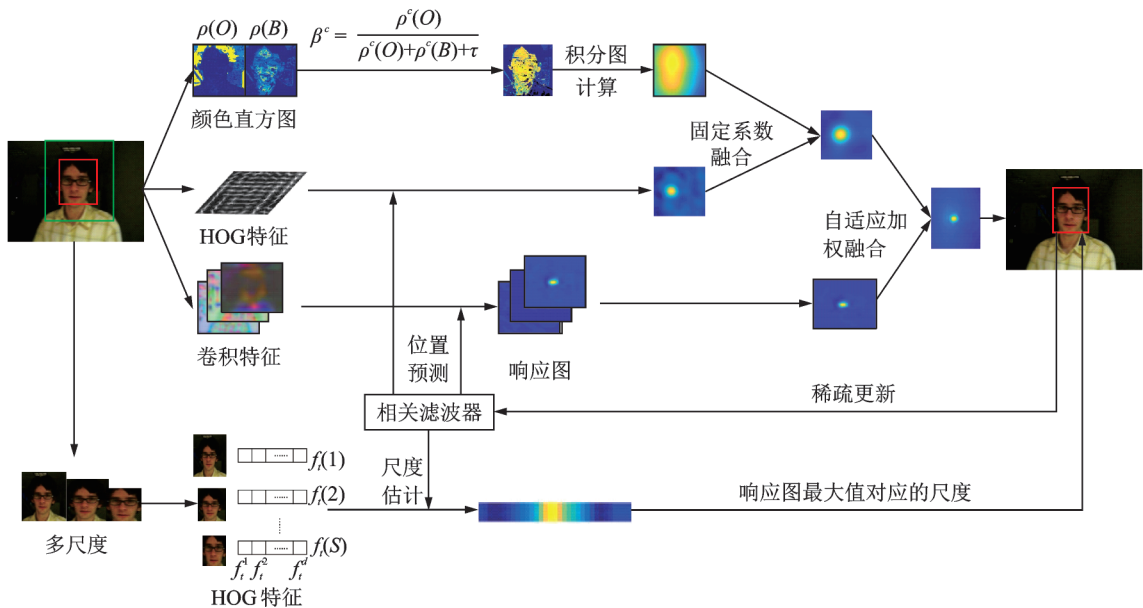


图4 算法的整体流程图

Fig.4 Flow chart of the proposed algorithm

步骤 1 滤波器训练

(1) 输入:第一帧图像 I_1 ;

(2) 在 I_1 的中心位置确定搜索区域图像,然后分别提取 HOG 特征、卷积特征以及 $\rho(O)$ 和 $\rho(B)$ 特征;

(3) 利用式(1)分别训练滤波器 W^{hog} , W^{conv} 和 W^{scale} 。

步骤 2 位置估计

(1) 输入:第 t 帧图像 I_t ,上一帧图像目标位置 P_{t-1} 和尺度 S_{t-1} ,滤波器 W_{t-1}^{hog} 和 W_{t-1}^{conv} , $\rho_{t-1}(O)$ 和 $\rho_{t-1}(B)$ 以及融合权重 k_{t-1} ;

(2) 在 I_t 位置 P_{t-1} 和尺度 S_{t-1} 上提取候选样本的 HOG 特征和卷积特征。利用相关滤波器 W_{t-1}^{hog} 通过式(3)计算得到 HOG 特征响应图 f_{hog} 。利用相关滤波器 W_{t-1}^{conv} 通过式(3,8)计算得到卷积特征响应图 f_{conv} ;

(3) 利用 $\rho_{t-1}(O)$ 和 $\rho_{t-1}(B)$,通过式(6)和积分图技术计算得到颜色直方图特征响应图 f_{color} ;

(4) 利用特征响应图 f_{hog} 和 f_{color} ,通过式(9)计算得到传统特征响应图 f_{trad} ;

(5) 通过式(10,11)计算得到权重 k_t ,然后利用权重对 f_{conv} 和 f_{trad} 进行自适应加权融合,得到最终响应图 f_{final} ,其最大值位置就是第 t 帧目标最终位置即为 P_t 。

步骤 3 尺度估计

(1) 输入:第 t 帧图像 I_t 及目标位置 P_t ,尺度相关滤波器 W_{t-1}^{scale} ;

(2) 在 I_t 位置 P_t 处提取 HOG 特征;

(3) 利用尺度相关滤波器 W_{t-1}^{scale} 和式(3)计算尺度响应图,其最大值位置就是第 t 帧目标估计尺度 S_t 。

步骤 4 模型更新

(1) 在 I_t 位置 P_t 和尺度 S_t 上提取 HOG 特征、卷积特征以及 $\rho_t(O)$ 和 $\rho_t(B)$ 特征,利用式(2)分别计算滤波器 W_t^{hog} , W_t^{conv} 和 W_t^{scale} ;

(2) 通过式(13)计算最终响应图 f_{final} 的 APCE 值,当 APCE 值大于更新阈值时,分别通过式(4),式(7)和式(12)对滤波器 W_t^{hog} , W_t^{conv} 和 W_t^{scale} , $\rho_t(O)$ 和 $\rho_t(B)$ 以及权重 k_t 进行在线更新;

(3) 将滤波器 W_t^{hog} , W_t^{conv} 和 W_t^{scale} , $\rho_t(O)$ 和 $\rho_t(B)$ 以及权重 k_t 传递至下一帧。

3 实验结果与分析**3.1 实验环境及参数设置**

本文所有的实验均在配置为 Intel Core i7-7800X 3.50 GHz CPU, GTX Titan X GPU, 内存为 16 GB 的台式电脑上完成,实验算法开发平台为 Matlab R2017b,卷积特征计算框架为 MatConvNet^[20]。实验中对所有的测试视频采用一样的参数,具体设置为:相关滤波器学习率 η_{corr} 为 0.01,颜色直方图滤波器学习率 η_{color} 为 0.04,权重学习率 η_k 为 0.1,正则化系数 λ 为 10^{-4} ,融合权重 k^{trad} 和 k^{conv} 初始值为 0.5,模型更新阈值为 30,固定权重 r 为 0.3,其值与 Staple 算法设置相同,网络卷积层的选取及每层的权重系数设置与 CF2 算法相同,尺度估计方法采用的参数与 DSST 算法设置一致。

3.2 数据集及评价指标

实验中选取 OTB-2013^[21] 标准数据集对算法进行评估,该数据集总共包含 51 个不同类型的视频序列。在实验结果的评估方面^[22],本文采用距离精度(Distance precision, DP)、重叠精度(Overlap precision, OP)以及一次性通过的精度曲线图(Precision plot)和成功率曲线图(Success plot)对算法进行评

估。DP定义为预测目标框中心与真实目标框中心之间的欧氏距离误差小于某一阈值的帧数占有所有帧数的百分比;OP定义为重叠率大于某一阈值的帧数占有所有帧数的百分比,其中重叠率的计算公式为 $(R_i \cap R_a)/(R_i \cup R_a)$, R_i 表示预测目标框区域, R_a 表示真实目标框区域。距离精度曲线图是距离精度值与阈值之间关系的曲线图,选取阈值为20像素时的距离精度值作为评估值;成功率曲线图是重叠精度值与阈值之间关系的曲线图,将成功率曲线下方的面积(Area under the curve, AUC)作为评估值。

为验证本文算法的有效性,总共设计了3组实验,具体如下。

实验1 多特征融合对比实验

选择本文算法所用特征进行拆分,得到两种算法 Ours_cnn 和 Ours_staple,其中 Ours_cnn 只利用卷积特征进行目标跟踪,Ours_staple 算法利用 HOG 和颜色直方图特征进行目标跟踪。为了充分证明本文多特征融合的有效性,分别选取 Ours_cnn, Ours_staple 以及其他融合不同特征的跟踪算法分别在 OTB-2013 测试集上与本文算法作对比分析试验。图5显示了各种跟踪算法的精度曲线图和成功率曲线图。

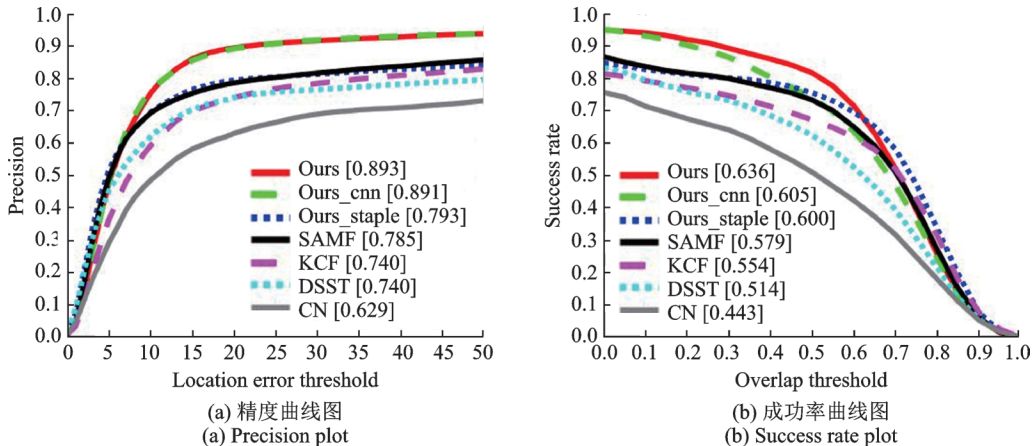


图5 7种算法的精度曲线图和成功率曲线图

Fig.5 Precision plots and success rate plots of seven algorithms

从图5可以看出,相比于使用HOG特征的KCF、使用颜色属性特征的CN、融合HOG和颜色属性特征的SAMF、融合HOG特征和灰度特征的DSST以及Ours_cnn和Ours_staple,本文算法的距离精度值和成功率在OTB-2013测试集上都位居第一。

实验2 定量对比实验

利用OTB-2013测试集对本文算法和主流的9种跟踪算法(CF2, HDT, CNN-SVM^[23], FCNT, ACFN^[24], SiamFC, SRDCF, DSST和SAMF)作一次性通过评估(One-pass evaluation, OPE)。为了充分说明本文算法的有效性,实验选取的跟踪算法非常具有针对性,其中,DSST和SAMF算法是利用传统特征并引入尺度估计进行目标跟踪的算法,算法简单高效;SRDCF算法是针对相关滤波器引入空间正则化的改进算法,算法复杂度高;CF2, HDT, CNN-SVM, FCNT, ACFN和SiamFC算法都是近年来基于深度学习,利用卷积特征学习的跟踪算法,算法运行速度较慢。图6为10种算法的精度曲线图和成功率曲线图。

从图6可以看出,本文算法在OTB-2013测试集上平均DP为89.3%,平均OP为63.6%。虽然本文算法的平均DP与其次优的CF2算法相比只提高了0.2%,但是平均OP却比CF2算法提高了3.1%。同

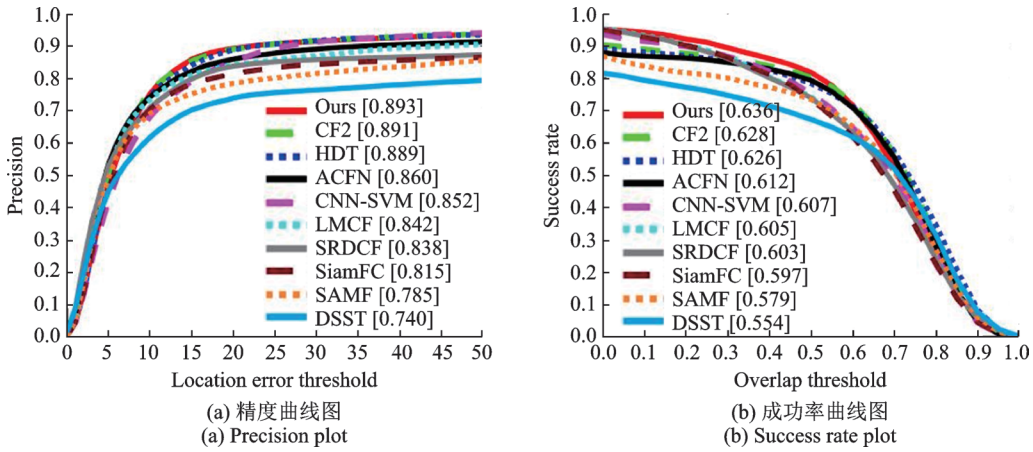


图6 10种算法的精度曲线图和成功率曲线图

Fig.6 Precision plots and success rate plots of ten algorithms

样,虽然本文算法的平均OP与其次优的SRDCF算法相比只提高了1.0%,但是平均DP却比SRDCF算法提高了5.5%。

表1分别为不同算法在不同挑战因素下的距离精度值,其中包含的挑战因素有遮挡(Occlusion, OCC)、光照变化(Illumination variation, IV)、尺度变化(Scale variation, SV)、快速移动(Fast motion, FM)、背景干扰(Background clutters, BC)、运动模糊(Motion blur, MB)、非刚性形变(Deformation, DEF)、平面内旋转(In-plane rotation, IPR)和平面外旋转(Out-of-plane rotation, OPR)。从表中数据可以看出,本文算法在尺度变化、遮挡、背景杂乱、运动模糊、平面内旋转和平面外旋转6种挑战因素中位居第一。对于目标发生非刚性形变的情况,本文算法、CF2和HDT算法都表现较好,但是FCNT算法着重分析了不同特征对目标定位的不同作用,对目标形变有很强的鲁棒性。此外,本文算法因颜色特征容易受光照变化的影响,故对光照变化的情况表现略显劣势。

表1 不同算法在不同挑战因素下的距离精度值

Tab. 1 Distance precisions of different algorithms on different attributes

算法	Ours	CF2	HDT	CNN-SVM	ACFN	FCNT	SiamFC	SRDCF	DSST	SAMF
FM	<u>0.789</u>	0.790	0.782	0.723	0.719	0.767	0.743	0.741	0.513	0.608
BC	0.886	<u>0.885</u>	0.871	0.789	0.783	0.799	0.742	0.803	0.694	0.676
MB	0.845	<u>0.844</u>	0.840	0.745	0.709	0.789	0.726	0.789	0.544	0.5640
DEF	0.885	0.881	0.884	0.885	<u>0.902</u>	0.917	0.730	0.855	0.658	0.810
IV	0.843	<u>0.844</u>	0.845	0.780	0.793	0.830	0.717	0.761	0.730	0.682
IPR	0.872	0.868	<u>0.869</u>	0.836	0.814	0.811	0.760	0.766	0.768	0.714
OCC	0.880	<u>0.877</u>	0.874	0.770	0.856	0.797	0.797	0.844	0.706	0.839
OPR	0.874	0.869	<u>0.871</u>	0.832	0.870	0.831	0.788	0.818	0.736	0.767
SV	0.886	<u>0.880</u>	0.866	0.827	0.813	0.830	0.802	0.778	0.738	0.723

注:表中排名第一的均用加粗体表示,排名第二的均用下划线表示。

实验3 定性对比实验

选取5种算法在6组具有各种挑战因素的视频序列的跟踪结果进行显示,定性对比分析本文算法在复杂环境下的跟踪性能。从实验2中可以看出本文算法性能远高于DSST, SAMF, SRDCF, CNN-SVM和SiamFC 5种算法,因此本实验中仅与剩余的4种算法进行比较。视频序列分别为Dog1, MotorRolling, Skating1, Soccer, Fleetface和SUV,其相关属性如表2所示。

表2 测试视频的属性

Tab. 2 Video sequences attributes

测试视频	帧数	目标大小	面临的主要挑战
Dog1	1 350	51×36	SV, IPR, OPR
MotorRolling	164	122×125	IV, SV, MB, FM, IPR, BC, LR
Skating1	400	34×84	IV, SV, OCC, DEF, OPR, BC
Soccer	393	67×81	IV, SV, OCC, MB, FM, IPR, OPR, BC
Fleetface	707	122×148	SV, DEF, MB, FM, IPR, OPR
SUV	945	91×40	OCC, IPR, OV

本文算法与4种算法在典型视频部分帧的跟踪结果如图7所示。图7(a)Dog1视频中,目标发生了巨大尺度变化,虽然各种算法都能跟踪到目标,但是只有本文算法对目标的尺度预测较为准确;图7(b)MotorRolling视频中,背景杂乱且目标发生了旋转,本文算法更能准确地跟踪到目标;图7(c)Skating1视频中,当目标遇到部分遮挡时,本文算法比其他算法更能准确跟踪目标,随后目标受到光照变化和尺度变化的影响,只有本文算法能够稳定地跟踪到目标;图7(d)Soccer视频中,目标遇到部分遮挡、运动模糊和背景杂乱等一系列干扰,只有本文算法能够准确追踪到目标。图7(e)Fleetface视频中,目标发生了尺度变化且在场景内发生了旋转,本文算法更准确地跟踪到目标。图7(f)SUV视频中,目标分辨率很低且超出了视野范围内或目标遇到完全遮挡,本文算法也能够鲁棒地跟踪目标。综上所述,本文算法在跟踪过程中遇到遮挡、尺度变化、形状变化和旋转等各种挑战时,具有较好的鲁棒性。

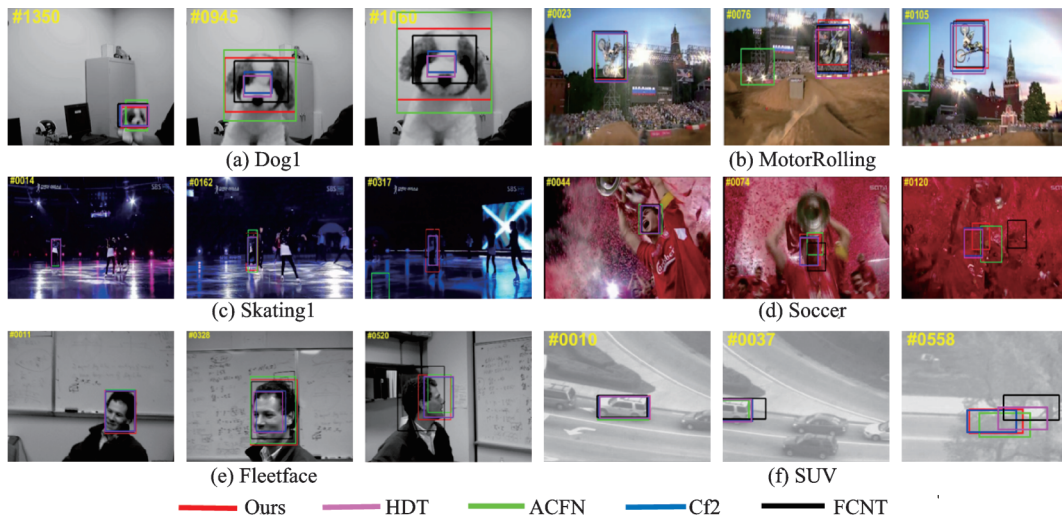


图7 5种算法在6个视频序列的跟踪结果比较

Fig.7 Visualization of tracking results of five algorithms on six sequences (Ours HDT ACFN Cf2 FCNT)

4 结束语

本文在相关滤波的基础上,提出了一种自适应融合传统特征(HOG特征与颜色直方图特征)和卷积特征的运动目标跟踪算法。该算法首先利用固定系数方法融合HOG和颜色直方图的特征响应图,然后将融合结果与卷积特征响应图通过自适应权重融合策略进行融合,最后基于融合后的响应图估计目标位置;采用OTB-2013公开测试集验证本文算法的性能,并与几种主流目标跟踪算法进行了对比分析。定量和定性实验结果表明,本文算法的准确性和鲁棒性都优于其他算法,由于多特征融合使用以及卷积特征的提取比较耗时,尽管稀疏更新策略使得算法跟踪速度有所提升,但本文所提出的算法速度仍然很慢。对此问题,将来进一步的研究工作可以考虑在卷积特征通道上进行分析,实现自适应多通道特征的选择,减少冗余特征的影响,从而提升特征提取的时效性,有利于改善跟踪算法的实时性和准确性。

参考文献:

- [1] 石祥滨,张健,代钦,等.采用显著性分割与目标检测的形变目标跟踪方法[J].计算机辅助设计与图形学学报,2016,28(4):645-653.
Shi Xiangbin, Zhang Jian, Dai Qin, et al. A deformed object tracking method utilizing saliency segmentation and target detection [J]. Journal of Computer-Aided Design and Computer Graphics, 2016, 28(4):645-653.
- [2] Smeulders A W M, Chu D M, Cucchiara R, et al. Visual tracking: An experimental survey[J]. The IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7):1442-1468.
- [3] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation on filters[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles, CA, USA: IEEE, 2010:2544-2550.
- [4] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters[J]. The IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(3):583-596.
- [5] Danelljan M, Häger G, Khan F S, et al. Accurate scale estimation for robust visual tracking[C]// Proceedings of British Machine Vision Conference. Nottingham, UK: BMVA Press, 2014:1-11.
- [6] Danelljan M, Khan F S, Felsberg M, et al. Adaptive color attributes for real-time visual tracking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA: IEEE, 2014:1090-1097.
- [7] Li Yang, Zhu Jianke. A scale adaptive kernel correlation filter tracker with feature integration[C]// Proceedings of the 13th Europe Conference on Computer Vision. Zurich, Switzerland: Springer, 2014:254-265.
- [8] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: Complementary learners for real-time tracking[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016:1401-1409.
- [9] Wang Naiyan, Yeung D Y. Learning a deep compact image representation for visual tracking[C]// Proceedings of Advances in Neural Information Processing Systems. Nevada, USA: MIT Press, 2013:809-817.
- [10] Ma Chao, Huang Jiabin, Yang Xiaokang, et al. Hierarchical convolutional features for visual tracking[C]// Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:3074-3082.
- [11] Qi Yuankai, Zhang Shengping, Qin Lei, et al. Hedged deep tracking[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016:4303-4311.
- [12] Wang Lijun, Ouyang Wanli, Wang Xiaogang, et al. Visual tracking with fully convolutional networks[C]// Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015:3119-3127.
- [13] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional siamese networks for object tracking[C]// Proceedings of the 14th Europe Conference on Computer Vision. Amsterdam, Netherlands: Springer, 2016:850-865.
- [14] Danelljan M, Häger G, Khan F, et al. Adaptive decontamination of the training set: A unified formulation for discriminative visual tracking[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016:1430-1438.
- [15] Danelljan M, Robinson A, Khan F, et al. Beyond correlation filters: Learning continuous convolution operators for visual tracking

- [C]//Proceedings of the 14th Europe Conference on Computer Vision. Amsterdam, Netherlands:Springer, 2016:472-488.
- [16] 赵高鹏,沈玉鹏,王建宇.基于核循环结构的自适应特征融合目标跟踪[J].光学学报,2017,37(8):208-217.
Zhao Gaopeng, Shen Yupeng, Wang Jianyu. Adaptive feature fusion object tracking based on circulant structure with kernel[J]. Acta Optica Sinica, 2017, 37(8):208-217.
- [17] Wang Mengmeng, Liu Yong, Huang Zeyi. Large margin object tracking with circulant feature maps[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition.Honolulu, Hawaii, USA:IEEE, 2017:4800-4808.
- [18] 杨丽娟,张白桦,叶旭楨.快速傅里叶变换FFT及其应用[J].光电工程,2004,31(S1):1-3.
Yang Lijuan, Zhang Baihua, Ye Xuzhen. Fast Fourier transform FFT and its application[J].Opto-Electronic Engineering, 2004, 31(S1):1-3.
- [19] 黄文杰,陈斌.一种快速图像处理的积分图方法[J].计算机应用,2005(S1):266-268.
Huang Wenjie, Chen Bin. An integral graph method for fast image processing[J].Journal of Computer Applications, 2005(S1): 266-268.
- [20] Vedaldi A, Lenc K. MatConvNet: Convolutional neural networks for Matlab[C]//ACM International Conference on Multimedia. Brisbane, Australia: ACM, 2015:689-692.
- [21] Wu Yi, Lim J, Yang M H. Online object tracking: A benchmark[C]// Proceedings of 2013 IEEE Conference on Computer Vision and Pattern Recognition. Portland, OR, USA:IEEE, 2013:2411-2418.
- [22] Everingham M, Van G L, Williams C K I, et al. The pascal visual object classes challenge[J].International Journal of Computer Vision, 2010, 88(2):303-338.
- [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 Machine Learning. Lille, France:JMLR, 2015:597-606.
- [24] Choi J, Chang H J, Yun S, et al. Attentional correlation filter network for adaptive visual tracking[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, Hawaii, USA:IEEE, 2017:4828-4837.

作者简介:



谢柳(1994-),男,硕士研究生,研究方向:计算机视觉与图像处理, E-mail: xie-liu4087@126.com。



尚振宏(1975-),男,通信作者,副教授,研究生导师,研究方向:计算机视觉与图像处理, E-mail: shang-zhenhong@126.com。



刘辉(1969-),男,教授,硕士生导师,研究方向:计算机视觉与模式识别, E-mail: liuhui2169@aliyun.com。

(编辑:王静)