

基于时序一致和空间剪裁的多特征相关滤波跟踪算法

王译萱, 吴小俊*, 徐天阳

江南大学物联网工程学院模式识别与计算智能国际联合实验室, 江苏 无锡 214122

摘要 为提升相关滤波跟踪算法在目标遮挡、背景嘈杂及目标形变等干扰下的跟踪精度, 提出一种基于时序一致和空间剪裁的多特征相关滤波跟踪算法。在训练阶段利用二值矩阵掩模对滤波器模板的能量分布进行裁剪, 使模板信息更加集中于目标区域, 从而缓解循环样本造成的边界效应; 利用 l_2 范数作为时序一致模型对相邻帧的滤波器建立平滑性约束, 使滤波器模板学习到相邻帧目标的上下文信息, 增加算法的抗干扰能力; 为进一步提升目标模板的表达能力, 将包含丰富语义信息的 ResNet50 深度特征引入到跟踪框架中, 通过主成分分析法对提取到的深度特征进行降维, 采用传统特征结合深度特征的方式提升跟踪结果的精确度和稳健性。将本文算法与 5 种算法进行对比实验, 验证了本文算法在处理目标遮挡、背景嘈杂及目标形变等干扰时的稳健性。

关键词 机器视觉; 相关滤波器; 时空一致模型; 二值矩阵掩模; 边界效应; 背景干扰

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.221502

Tracking Algorithm of Correlation Filter with Multiple Features Based on Temporal Consistency and Spatial Pruning

Wang Yixuan, Wu Xiaojun*, Xu Tianyang

International Joint Laboratory of Pattern Recognition and Computational Intelligence, School of Internet of Things Engineering, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract Aiming to improve the tracking accuracy of the correlation filter tracking algorithm when faced with occlusion, background clutter, and deformation of the object target, this study proposes a correlation filter tracking algorithm for multiple features based on temporal consistency and spatial pruning. First, in the training stage, the energy distribution of the filter template is pruned using the binary matrix mask to make the template information more concentrated in the target area, which alleviates the boundary effect caused by the cyclic shifted samples. Second, the l_2 -norm is used as the temporal consistency model to establish smoothness constraints for the filters of two consecutive frames so that filter templates can learn the context information of consecutive-frame target and increase the anti-interference ability of the algorithm. To further improve the expressive ability of the target template, ResNet50 deep features, which contain rich semantic information, are introduced into the tracking framework. Principal component analysis is used to reduce the dimension of the extracted depth feature, and traditional features in combination with deep features improve the accuracy and robustness of the tracking results. A comparison of the proposed algorithm with five existing algorithms verifies the proposed tracking algorithm's robustness in dealing with distractors such as target occlusion, background clutter, and deformation.

Key words machine vision; correlation filter; temporal consistency model; binary matrix mask; boundary effects; background interference

OCIS codes 150.1135; 100.4999; 070.4550

1 引言

目标跟踪是计算机视觉中的重要研究课题之一,

在视频监控、无人驾驶和人机交互等方向有着广泛的应用。跟踪的主要目的是对视频序列中的运动目标进行检测、提取、识别和跟踪, 获得运动目标的表现信

收稿日期: 2019-04-11; 修回日期: 2019-04-29; 录用日期: 2019-05-09

基金项目: 国家自然科学基金(61672265, U1836218)、高等学校学科创新引智计划(B12018)

* E-mail: xiaojun_wu_jnu@163.com

息和运动参数,从而进行相应的处理与分析^[1]。

近些年,相关滤波跟踪算法因拥有较高的跟踪精度和实时性等优势,在众多公开视频测试数据集和竞赛中取得了优异的成绩,因而受到国内外研究人员的广泛关注^[2]。影响相关滤波跟踪算法性能的因素主要包括跟踪框架的设计、表现信息的特征提取方式、选择和融合等^[3]。

早期的相关滤波跟踪算法由于在训练过程中缺少训练样本,导致跟踪精度较差,直到2010年Bolme等^[4]提出输出误差最小平方和滤波器(MOSSE),跟踪效果才得以改善。Henriques等^[5-6]在此基础上,提出循环结构跟踪算法(CSK)和核相关滤波跟踪算法(KCF),利用循环矩阵结构生成大量的循环样本,有效提升了分类器的鉴别能力,获得较好的跟踪精度。此外,研究人员针对目标存在的尺度变化问题,将鉴别性尺度滤波器引入跟踪框架中^[7-8],以构建尺度回归和尺度采样的方法对目标的尺度变化进行准确的估计,使得算法的跟踪精度有较大的提升。近年来,更多相关滤波算法集中于优化由循环矩阵结构导致的边界效应问题。循环矩阵结构可以有效解决跟踪训练过程中训练样本较少的问题,但经由循环结构生成的循环样本往往包含着较多的边界信息,过多的高频信息会对滤波器的训练造成较大的干扰,导致分类器的鉴别能力下降。针对这一问题的改进算法有空间正则化相关滤波器(SRDCF)^[9]和空间通道可靠性滤波器(CSRDCF)^[10]等。其中,SRDCF在目标函数中对滤波器的空间分布设置空间正则化约束,使滤波器的能量主要集中于目标区域,从而达到在抑制边界效应的同时,扩大目标的搜索区域,提升算法精度的目的;CSRDCF则利用目标图像的颜色直方图特征判断前景与背景区域,以此作为依据生成掩模,将目标前景与背景分离,有效地去除了循环矩阵结构引入的混叠样本,提升了训练样本的质量。

通过对上述算法的研究,本文提出一种基于时序一致和空间剪裁的多特征相关滤波算法。首先,扩大算法中的目标采样区域,获取更多的循环样本,同时利用学习得到的二值矩阵对滤波器模板周围的能量进行空间裁剪,使模板信息集中于目标区域,抑制循环样本产生的边界效应,提升样本质量;其次,利用 l_2 范数作为时序一致模型,使滤波器模板学习到目标相邻帧的上下文信息,缓解样本被污染时的滤波器退化问题,增强算法的抗干扰能力;最后引入ResNet50深度特征增加目标模板的表达能,利用主成分分析法(PCA)对深度特征进行降维,降低高

维特征带来的复杂计算量。

2 多通道相关滤波跟踪算法原理

相关滤波算法以目标为中心,在目标周围循环采样,实现滤波器的训练,通过滤波器与候选目标的相关操作,判断目标所在位置^[5]。算法的跟踪过程主要包括滤波器训练、目标检测以及模板更新3部分^[3,8]。

2.1 滤波器训练

相关滤波跟踪算法通过训练滤波器,将跟踪目标与背景分离,其训练过程为标准的岭回归过程^[6],即

$$\epsilon = \sum_{d=1}^{N_d} \| \mathbf{f}_d * \mathbf{h}_d - \mathbf{g} \|^2 + \lambda \sum_{d=1}^{N_d} \| \mathbf{h}_d \|^2, \quad (1)$$

式中, ϵ 代表损失函数, \mathbf{h}_d 代表滤波器,*代表相关运算, \mathbf{f}_d 代表当前帧目标样本的第 d 个特征通道输入, N_d 代表特征通道总和, \mathbf{g} 代表样本标签, λ 代表正则项系数,防止过拟合。对(1)式进行快速傅里叶变换(FFT),得到滤波器在频域内的闭环解,即

$$\hat{\mathbf{h}}_d = \frac{\hat{\mathbf{g}}^H \odot \hat{\mathbf{f}}_d}{\sum_{d=1}^{N_d} \hat{\mathbf{f}}_d^H \odot \hat{\mathbf{f}}_d + \lambda}, \quad (2)$$

式中, $\hat{\mathbf{f}}_d$ 和 $\hat{\mathbf{h}}_d$ 分别为 \mathbf{f}_d 和 \mathbf{h}_d 的傅里叶变换,H为厄米特转置, \odot 为频率域中的点乘^[5-6]。(2)式即为训练得到的滤波器。

2.2 目标检测

利用滤波器检测下一帧目标的位置,由于相邻帧时间间隔较短,目标通常不会大范围移动,因此需根据目标上一帧的位置建立采样区域,对采集测试样本 \mathbf{z} 与分类器 $\hat{\mathbf{h}}$ 进行点乘操作,得到目标响应,响应峰值区域即为目标位置,具体表达式为

$$\mathbf{R}(\mathbf{z}) = \mathcal{F}^{-1} \{ \hat{\mathbf{h}} \odot \hat{\mathbf{z}} \}, \quad (3)$$

式中, $\mathbf{R}(\mathbf{z})$ 表示当前帧的目标响应值, \mathcal{F}^{-1} 表示快速傅里叶逆变换(IFFT)。

2.3 模板更新

考虑到目标在跟踪过程中的外观变化,需要对学习到的滤波器进行在线更新,因此在第 t 帧视频序列中相关滤波器 $\hat{\mathbf{h}}$ 的更新公式为

$$\hat{\mathbf{h}} = (1 - \eta) \hat{\mathbf{h}}_{t-1} + \eta \hat{\mathbf{h}}_t, \quad (4)$$

式中, η 表示学习速率,作为滤波器模板的更新参数, t 表示当前帧数。

2.4 问题分析

传统相关滤波跟踪算法需要大量训练样本训练滤波器,但在实际的跟踪过程中,仅在第1帧指定真实的跟踪目标,样本数量不足。为提升样本数量,多数算法对目标周围建立一个采样区域,以区域内的图像作为原始样本,在空间域内采用循环移位的方式生成一定数量的循环样本^[5],采样区域越大,获取的样本数量越多,但该方式产生的边界效应一定程度上降低了滤波器的鉴别能力,因此多数算法只能将采样区域控制在相对较小的范围内,通常为目标大小的1.5至3倍。CSRDCF针对该问题提出基于颜色直方图特征的前景掩模,通过前景掩模抑制滤波器模板周围的能量,降低边界效应对跟踪造成的影响,提升了算法的跟踪精度。但该方法获取掩模的过程依然存在如下缺点:1)颜色直方图特征稳健性较弱,容易受到背景及光照变化的影响,导致分割结果不够准确;2)当背景与前景相似程度较高时,分割后的前景掩模可能包含目标的背景信息,导致滤波器边缘区域依然存在能量;3)获取颜色直方图特征的过程增加了算法的额外耗时;4)分割得到的前景掩模空间范围较小,不利于提取表达能力更强但分辨率较低深度特征。

此外,多数相关滤波跟踪算法在特征提取阶段采用方向梯度直方图特征(HOG)结合颜色名称特征(CN)的形式^[10-11],这种传统特征组合具有较高的分辨率,提取过程相对简单,在跟踪速度上具有一定优势,但在目标发生旋转、快速运动及复杂背景下的稳健性较差,不利于算法实现长时稳健跟踪。

3 相关滤波算法的改进

为解决上述问题,在CSRDCF的基础上进行改进,提出一种基于时序一致和空间剪裁的多特征相关滤波跟踪算法,算法的完整流程图如图1所示,改进部分用加粗的方式在图中进行标注。首先在初始帧的滤波器训练过程中加入二值矩阵掩模约束,通过二值矩阵对滤波器的模板信息(目标-非目标)进行裁剪,以达到抑制边界效应的目的。在后续的滤波器训练过程中,加入基于 l_2 范数的时序一致模型,对目标表现在时序上的平滑性进行建模,使滤波器模板学习到目标的上下文信息,增强算法的抗干扰能力。采用固定迭代次数的增广拉格朗日乘法(Augmented Lagrange Method, ALM)^[12]对滤波器模板进行求解。此外,针对CSRDCF算法中使用的HOG和CN两种

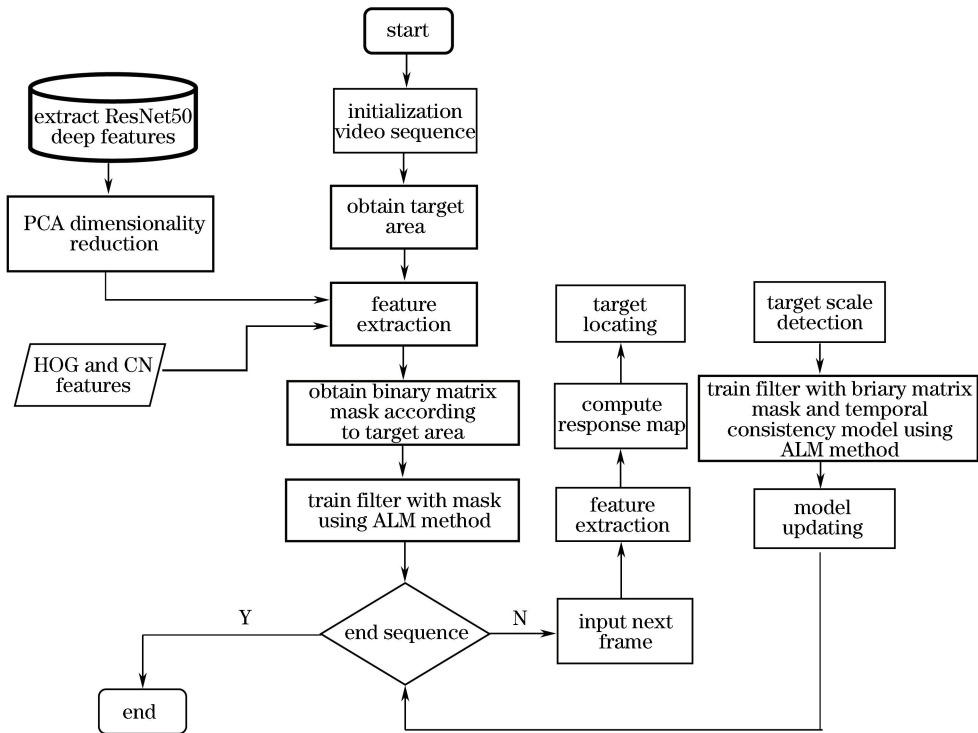


图1 基于时序一致和空间剪裁的多特征相关滤波跟踪算法流程图

Fig. 1 Flow chart of correlation filter tracking algorithm for multiple features based on temporal consistency and spatial pruning

传统特征在复杂背景下的稳健性不足问题,本文在特征提取阶段加入了 PCA 降维后的 ResNet50 深度特征,提升了算法整体的稳健性。

3.1 二值矩阵掩模

为解决 CSRDCF 中前景掩模存在的不足,提出二值矩阵掩模,其作用与前景掩模类似。利用二值矩阵实现滤波器能量的裁剪,即

$$h_m = m \odot h, \quad (5)$$

式中 m 为二值矩阵掩模,矩阵中只包含 0 或 1 元素,二值矩阵的维度与目标区域大小相同; h 表示滤波器模板; h_m 表示 h 与 m 点乘后的滤波器。通过二值矩阵掩模 m 实现滤波器模板能量裁剪,只保留模板在目标区域内的信息,达到抑制边界效应的目的,其示意图如图 2 所示,其中 x_{t-1} 为第 $t-1$ 帧中用于训练滤波器的目标图像块, x_t 为第 t 帧中用于训练滤波器的目标图像块, y 为高斯形状的期望输出。

区别于同样使用掩模的 CSRDCF,本文提出的

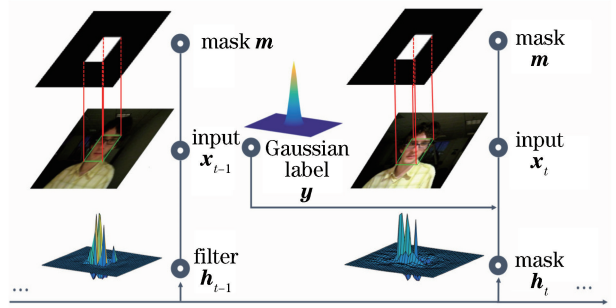


图 2 二值矩阵掩模示意图

Fig. 2 Schematic of binary matrix mask

掩模操作无需获取目标的颜色直方图特征,所以耗时更短,同时避免了颜色特征稳健性不足的缺陷。此外,二值矩阵掩模相比于 CSRDCF 中的前景掩模具有更大的空间范围。因此,在使用深度特征时,本文方法可以学习到更为丰富的语义信息,使目标模板具有更强的表达能力,两种算法得到的滤波器模板对比图如图 3 所示。

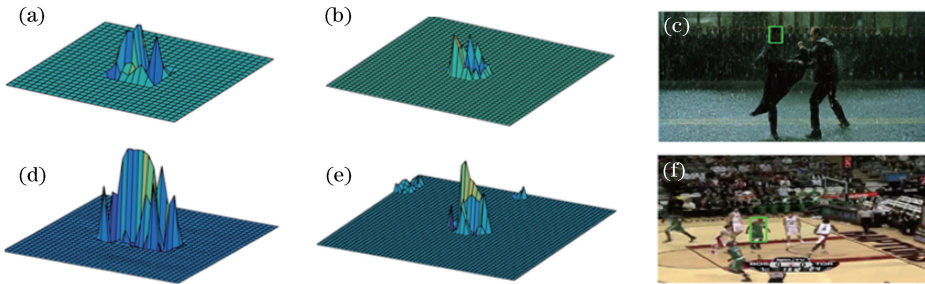


图 3 本文算法与 CSRDCF 生成的掩模对比图。(a)本文算法基于 *Matrix* 视频中单帧图像生成的掩模;(b) CSRDCF 算法基于 *Matrix* 视频中单帧图像生成的掩模;(c) *Matrix* 视频中的单帧图像;(d)本文算法基于 *Basketball* 视频中单帧图像生成的掩模;(e) CSRDCF 算法基于 *Basketball* 视频中单帧图像生成的掩模;(f) *Basketball* 视频中的单帧图像

Fig. 3 Comparison of generated masks by proposed algorithm and CSRDCF algorithm. (a) Mask generated by proposed algorithm based on by single frame image in *Matrix* video; (b) mask generated by CSRDCF algorithm based on by single frame image in *Matrix* video; (c) single frame image in *Matrix* video; (d) mask generated by proposed algorithm based on by single frame image in *Basketball* video; (e) mask generated by CSRDCF algorithm based on by single frame image in *Basketball* video; (f) single frame image in *Basketball* video

3.2 时序一致模型

相关滤波跟踪算法的目标模板在跟踪过程中需要不断更新,因此当跟踪目标受到遮挡、光照变化及背景嘈杂等干扰时,用于训练滤波器的目标样本可能会被污染,从而导致跟踪漂移或失败。针对该问题,本文在跟踪算法中加入对目标历史帧模板信息的考量,提升了跟踪结果的稳健性。通过实验发现,跟踪过程中,相邻帧的时间间隔较短,因此目标和背景在前后两帧间通常不会产生太大的差异。本文以此为依据,利用 l_2 范数作为时序一致模型对相邻帧的滤波器模板进行约束,使目标模板学习到目标的上下文信息,即

$$\min \| h_t - h_{t-1} \|^2, \quad (6)$$

式中 h_t 和 h_{t-1} 分别表示当前帧和上一帧的滤波器模板。(6)式使 h_t 和 h_{t-1} 的误差平方和最小,从时序上对目标模板进行相似性约束,使滤波器模板一定程度上保留历史帧的模板信息,即使当前帧目标受到干扰导致训练样本污染,学习到目标上下文信息的滤波器模板依然能够保持较强的鉴别能力,从而降低退化风险。为方便优化表示,后文用 h_p 替代 h_{t-1} 表示上一帧的滤波器模板。

3.3 增广拉格朗日优化过程

将(5)式与(6)式作为约束条件加入(1)式中,即

$$\epsilon = \sum_{d=1}^{N_d} \| \mathbf{f}_d * \mathbf{h}_d - \mathbf{g} \|^2 + \lambda_1 \sum_{d=1}^{N_d} \| \mathbf{m} \odot \mathbf{h}_d \|^2 + \lambda_2 \sum_{d=1}^{N_d} \| \mathbf{h}_d - \mathbf{h}_p \|^2, \quad (7)$$

式中, λ_1, λ_2 为正则化约束项的权重。(7)式中的变量 \mathbf{h}_d 在频域内没有闭环解,因此采用固定迭代次数的 ALM 方法对(7)式进行优化求解,优化过程与 CSRDCF 类似,引入松弛变量 $\mathbf{h}'_d = \mathbf{h}_d$,交替优化变量 \mathbf{h}'_d 和 \mathbf{h}_d 。将(7)式进行拉格朗日优化变形,即

$$l = \sum_{d=1}^{N_d} \| \mathbf{f}_d * \mathbf{h}_d - \mathbf{g} \|^2 + \lambda_1 \sum_{d=1}^{N_d} \| \mathbf{m} \odot \mathbf{h}'_d \|^2 + \lambda_2 \sum_{d=1}^{N_d} \| \mathbf{h}_d - \mathbf{h}_p \|^2 + \frac{\mu}{2} \sum_{d=1}^{N_d} \| \mathbf{h}_d - \mathbf{h}'_d + \frac{\mathbf{y}_d}{\mu} \|^2, \quad (8)$$

式中, l 表示拉格朗日优化式的损失, \mathbf{h}_p 表示上一帧优化得到的滤波器模板,优化过程中可作为常量处理。在第 1 帧的训练过程中, $\mathbf{h}_p = \mathbf{0}$, \mathbf{y}_d 为拉格朗日乘子, μ 为惩罚系数。对(8)式采用交替方向乘子法求解变量 \mathbf{h}'_d 与 $\hat{\mathbf{h}}_d$ ^[10,12],得到两个变量的闭环解,即

$$\hat{\mathbf{h}}_d = \frac{2\hat{\mathbf{f}}_d \odot \hat{\mathbf{g}}^H + \mu \hat{\mathbf{h}}'_d - \hat{\mathbf{y}}_d + 2\lambda_2 \hat{\mathbf{h}}_p}{\mu + 2\hat{\mathbf{f}}_d \odot \hat{\mathbf{f}}_d^H + 2\lambda_2}, \quad (9)$$

$$\mathbf{h}'_d = (\mathbf{1} - \mathbf{m}) \odot \frac{\mu \mathbf{h}_d + \mathbf{y}_d}{2\lambda_1 + \mu}, \quad (10)$$

式中, $\hat{\mathbf{h}}_p$ 为经过 FFT 的上一帧得到的滤波器模板, \mathbf{m} 为二值矩阵掩模。在滤波器训练的初始阶段, \mathbf{m} 可由给定目标大小和位置(Ground-truth)求得。对(9)式和(10)式进行迭代求解,随着迭代次数的增加,算法中的变量会逐渐收敛,迭代的过程中,需对拉格朗日乘子 \mathbf{y}_d 和惩罚系数 μ 更新,更新公式分别为 $\mathbf{y}_d = \mathbf{y}_d + \mu(\mathbf{h} - \mathbf{h}')$, $\mu = \min(\beta\mu, \mu_{\max})$,其中 β 为学习速率, μ_{\max} 为惩罚系数最大值。

3.4 深度特征与 PCA 降维

目标特征类型的选择在一定程度上影响着算法的跟踪结果,近年来多数相关滤波跟踪算法在特征类型的选择上以 HOG 融合 CN^[12-13] 为主,这种传统特征组合在跟踪速度上具有一定优势,但在目标发生旋转、快速运动及复杂背景下的稳健性较差。与传统特征相比,利用神经网络学习到的深度特征包含更加丰富的语义信息,跟踪中目标产生的旋转、平移和缩放具有不变性,在复杂背景下,具有更强的稳健性。因此本文算法采用了 HOG、CN 和利用 ImageNet 数据集预训练得到的 ResNet50 深度特征^[14]。

ResNet50 的网络结构允许保留之前网络层的一定比例输出,通过学习上一个网络的残差,减少信息传递的损耗和丢失,保护了信息的完整性,对描述目标外观具有很强的表达能力。本文根据文献[15]中针对 ResNet50 深度网络的测试结果,选择该网络中第 4 个卷积块(block-4)中第 5 个卷积层(conv5)的输出作为学习的深度特征,其输出为 1024 个特征图谱,属于网络中高层卷积特征,包含较多的语义信息,同时具有较强的稳健性,有利于本文算法提升跟踪精度。为减少高维特征带来的复杂计算量,本文采用文献[12]中的方法,利用 PCA 对得到的深度特征进行降维,从 1024 个特征图谱中提取其中的 K 个主要特征用于滤波器训练。该方法可以在保证跟踪精度的前提下,减少系统运算量,提升跟踪速度。

在目标检测的过程中,本文对 3 种特征采用文献[16]中的融合方式,对 HOG、CN 和 ResNet50 特征分别计算其对应的目标响应,将得到的结果缩放至同一分辨率下求和,即为最终的目标响应。

3.5 尺度估计

针对目标的尺度变化问题,本文采用鉴别性尺度滤波器(DSST)^[7]中的尺度估计法,根据估测出的目标位置,利用尺度滤波器在尺度空间进行搜索,以实现准确的目标尺度估计。

4 实验结果与分析

为评估本文算法的性能,利用数据集 OTB-100^[17]和 VOT2016^[18]对算法进行测试,算法的实现平台为 MATLAB R2016a。实验共分 2 部分:第 1 部分为实验定量分析,利用 OTB-100 和 VOT2016 数据集对本文算法进行总体性能对比测试,测试序列数量分别为 100 和 60;第 2 部分为实验定性分析,在 OTB-100 数据集中 4 组具有遮挡、背景杂乱及光照变化干扰属性的视频序列中,测试算法的跟踪效果。实验代码中 $\beta=1.5$,正则化系数 $\lambda_1=1, \lambda_2=5$,惩罚系数 $\mu=1$,惩罚系数最大值 $\mu_{\max}=10$,ALM 过程迭代 2 次,降维后的特征图谱 K 为 128。

4.1 数据集和评价指标

针对 OTB-100 数据集,采用一次通过评估模式(OPE),评估标准采用成功率曲线面积(AUC)和距离精度(DP)。AUC 为计算重叠率大于某个阈值的帧数占总帧数的百分比,其结果用曲线围成的面积表示;DP 为中心误差小于 20 pixel 的帧数占总帧数的百分比。

针对 VOT2016 数据集,算法评估标准采用期望

平均覆盖率(EAO)、失败率(Failures)、平均覆盖率(AO)3项指标。EAO综合衡量算法的精度和稳健性;Failures表示算法跟踪过程中丢失目标的次数,用来衡量算法的稳健性;AO表示算法的跟踪精度。

4.2 定量分析

4.2.1 OTB数据集实验结果

针对OTB数据集中100组具有复杂属性的视频序列,将算法在所有视频序列上运行一次,得到算法在100组视频序列中的平均AUC和DP,对比算法包括7种相关滤波算法,即本文采用ResNet50+HOG+CN特征组合的算法ours(D+HC)、本文采用HOG+CN特征组合的算法ours(HC)、ECO-HC(efficient convolution operators)^[16]、CCOT(continuous convolution operators for visual tracking)^[19]、BACF(background aware correlation filter)^[20]、SRDCF^[9]、CSRDCF^[10],实验结果如图4

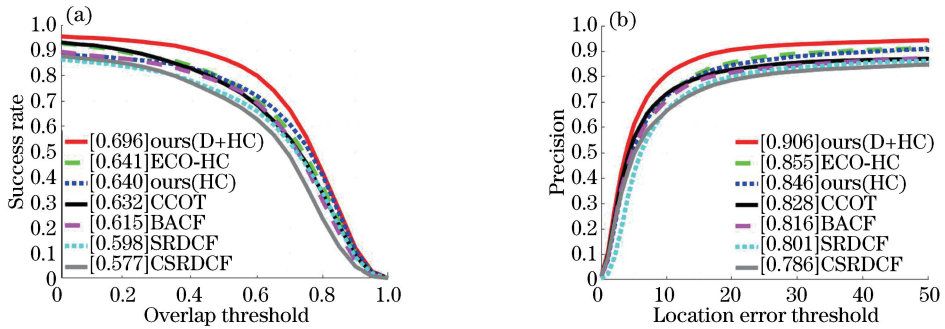


图4 7种算法在OTB-100上的平均AUC曲线和DP曲线。(a)平均AUC;(b)平均DP

Fig. 4 Average AUC and average DP curves of 7 algorithms on OTB-100 dataset. (a) Average AUC; (b) average DP

表1 6种算法在OTB-100上6种挑战中的平均成功率结果

Table 1 Average AUC results of 6 algorithms in 6 challenges on OTB-100 dataset

Algorithm	OCC	BC	DEF	MB	IPR	SV
Ours(D+HC)	0.669	0.694	0.651	0.699	0.658	0.670
ECO-HC	0.614	0.640	0.612	0.629	0.606	0.608
CCOT	0.620	0.613	0.585	0.663	0.582	0.601
BACF	0.556	0.605	0.572	0.570	0.575	0.575
SRDCF	0.554	0.583	0.540	0.590	0.542	0.565
CSRDCF	0.530	0.544	0.531	0.583	0.524	0.528

Ours(D+HC)在图4中的平均AUC值为69.6%,平均DP值为90.6%。Ours(D+HC)算法的AUC比ours(HC)算法提升了5.6%;比ECO-HC算法提升了5.5%;比CCOT算法提升了6.4%;比CSRDCF算法提升了11.9%。Ours(D+HC)算法的平均DP值比ours(HC)算法提升了7.8%;比ECO-HC算法提升了6%;比CCOT算法提升了5.1%;比CSRDCF算法提升了10.5%。此外,根据表1和表2中的数据可以看出,ours(D+HC)算法

所示。表1为ours(D+HC)与5种对比算法在遮挡(OCC)、背景杂乱(BC)、形变(DEF)、运动模糊(MB)、平面内旋转(IPR)、尺度变化(SV)6种挑战下的AUC和DP。

从图4中的实验对比结果可以看出,ours(HC)在100组视频序列中的平均AUC值为64.0%,低于ECO-HC算法0.1%;平均DP值为82.8%,低于CCOT算法2.7%,但相比于同类改进算法CSRDCF仍具有一定优势,平均AUC提升约6.3%,平均DP约提升2.7%。此外,本文算法的采样区域大小设置为目标的4.2倍,而CSRDCF算法的采样区域大小设置为目标的3倍,因此在滤波器的训练阶段,本文方法可以获取数量更多的循环样本。综合图4和表1中的实验结果,可以说明二值矩阵掩模和时序一致模型对于算法跟踪精度和稳健性的提升具有一定改进效果。

在遮挡、背景杂乱、形变、运动模糊、平面内旋转及尺度变化6种挑战中的AUC值和DP值均为最优,具有相对较高的稳健性,结合图4中的实验结果,充分说明在HOG和CN的基础上加入深度特征可以有效提升算法整体的跟踪精度和稳健性。

4.2.2 VOT2016数据集实验结果

为进一步验证所提算法的性能,利用数据集VOT2016对ours(D+HC)进行测试,该数据集中共包含70种跟踪算法。本节选取其中排名前8的

表 2 6 种算法在 OTB-100 上 6 种挑战中的平均精度结果

Table 2 Average DP results of 6 algorithms in 6 challenges on OTB-100 dataset

Algorithm	OCC	BC	DEF	MB	IPR	SV
Ours(D+HC)	0.869	0.914	0.870	0.888	0.882	0.882
ECO-HC	0.810	0.843	0.825	0.790	0.778	0.803
CCOT	0.851	0.832	0.828	0.844	0.833	0.822
BACF	0.731	0.801	0.766	0.733	0.790	0.773
SRDCF	0.728	0.775	0.731	0.760	0.738	0.747
CSRDCF	0.732	0.756	0.750	0.747	0.748	0.750

跟踪算法与本文算法进行对比测试,实验结果如

图 5和表 3 所示。

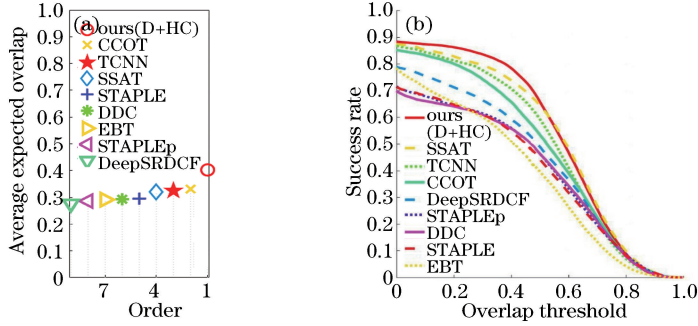


图 5 9 种算法在 VOT2016 数据集上的对比结果。(a) EAO;(b) AUC

Fig. 5 Comparison of 9 algorithms on VOT2016 dataset. (a) EAO; (b) AUC

表 3 9 种算法在 VOT2016 数据集上的跟踪结果

Table 3 Tracking results of 9 algorithms on VOT2016 dataset

Performance	Ours (D+HC)	CCOT	TCNN	SSAT	STAPLE	DDC	EBT	STAPLEp	DeepSRDCF
EAO	0.40	0.33	0.32	0.32	0.30	0.29	0.29	0.29	0.28
Failures	8.92	16.58	17.94	19.27	23.90	20.98	15.19	24.32	20.35
AO	0.53	0.47	0.49	0.52	0.39	0.39	0.37	0.39	0.43

从图 5 和表 3 显示的实验结果可以看出,本文算法的 EAO、Failures 和 AO 三项指标均明显优于其他跟踪算法,在 Failures 指标上比次优算法 CCOT 降低了约 50%,综合本文算法在 OTB-100 数据集上的实验结果,验证本文算法在跟踪精度和稳健性提升方面具有一定的改进效果。

4.3 定性分析

选取 4 组包含目标遮挡、形变和背景杂乱等干

扰信息的代表性视频序列,将 ours(D+HC)与 5 种算法进行对比测试,验证算法的稳健性。本文算法依然用 ours 表示,在视频中显示为红色矩形框。

4.3.1 目标遮挡性能分析

图 6 为 6 种算法在 Bird1 和 Lemming 两组视频上的跟踪效果图。在 Bird1 视频第 10 帧时,所有算法均可准确跟踪;在第 124 帧时目标开始发生遮挡,SRDCF

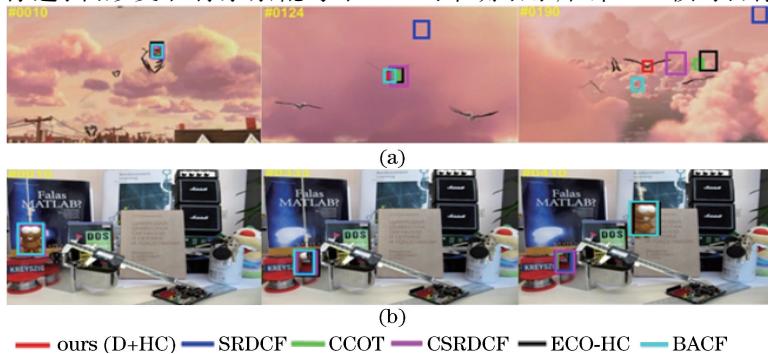


图 6 6 种算法在 Bird1 和 Lemming 上的跟踪效果图。(a) Bird1;(b) Lemming

Fig. 6 Comparison of tracking results of 6 algorithms on Bird1 and Lemming. (a) Bird1; (b) Lemming

丢失跟踪目标;在第 190 帧时目标经过长时间完全遮挡(约 60 帧)并重新出现,此时除本文算法外的其余算法均因目标模板污染丢失跟踪目标,而本文算法依然可以保持精确稳健跟踪。在 Lemming 数据集中,目标在第 333 帧时开始产生严重遮挡;在第 410 帧时目标重新出现,此时 ours(D+HC)、BACF 以及 ECO-HC 均可以准确跟踪目标。以上实验说明,本文算法能够较好地应对目标遮挡情况。

4.3.2 背景杂乱性能分析

图 7 为 6 种算法在 *Iron Man* 和 *Matrix* 两组

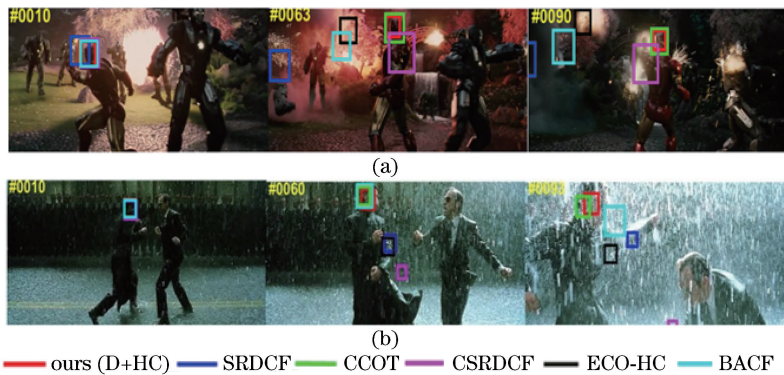


图 7 6 种算法在 *Iron Man* 和 *Matrix* 上的跟踪效果图。(a) *Iron Man*; (b) *Matrix*

Fig. 7 Comparison of tracking results of 6 algorithms on *Iron Man* and *Matrix*. (a) *Iron Man*; (b) *Matrix*

5 结 论

为提升相关滤波跟踪算法在复杂背景下的稳健性,同时降低边界效应对跟踪结果的影响,在 CSRDCF 的基础上,提出一种基于时序一致和空间剪裁的多特征相关滤波跟踪算法。算法在训练阶段利用空间范围更大的二值矩阵掩模对滤波器的能量分布进行裁剪,有效解决了采样窗口扩大导致的边界效应问题;基于 l_2 范数的时序一致模型使滤波器模板保留相邻帧目标的上下文信息,增强了算法在复杂背景下的抗干扰能力;ResNet50 深度特征弥补了 HOG 和 CN 稳健性不足的缺点,进一步提升了算法的跟踪精度。利用 OTB-2015 和 VOT2016 数据集对所提算法进行测试,验证了本文算法在面对遮挡、背景杂乱及目标形变等干扰时具有一定优越性。目前本文算法存在的主要缺点为运行速度较慢,其主要原因是运行深度网络和 ALM 迭代的过程增加了算法耗时。未来的工作中可以考虑对算法框架进行优化,以提升跟踪速度。

参 考 文 献

[1] Yang J F, Zhang J P. Long time target tracking

视频上的跟踪效果图。*Iron Man* 视频在第 10 帧初始化时背景较为杂乱,此时除 SRDCF 外,其余算法均能跟踪到目标;在第 63 帧和第 90 帧时背景干扰增强,目标发生形变和快速运动,此时只有本文算法和 CCOT 能够跟踪到目标,且本文算法的跟踪精度要优于 CCOT。在 *Matrix* 视频中,目标在第 10 帧、第 60 帧和第 93 帧之间发生了快速移动、形变、且背景与目标相似程度较高,在视频结束部分,尽管本文算法跟踪结果产生了偏差,但相较于其他算法仍然具有优势。

based on kernel correlation filtering [J]. *Laser & Optoelectronics Progress*, 2019, 56(2): 021502.

杨剑锋, 张建鹏. 基于核相关滤波的长时间目标跟踪 [J]. *激光与光电子学进展*, 2019, 56(2): 021502.

[2] Cheng Y, Li J Z, Zhu L N, *et al.* Correlation filter tracking algorithm based on model and scale updating [J]. *Laser & Optoelectronics Progress*, 2018, 55(12): 121015.

成悦, 李建增, 褚丽娜, 等. 基于模型与尺度更新的相关滤波跟踪算法 [J]. *激光与光电子学进展*, 2018, 55(12): 121015.

[3] Feng F, Wu X J, Xu T Y. Object tracking with multiple memory learning and adaptive correlation filter based on subspace and histogram [J]. *Pattern Recognition and Artificial Intelligence*, 2018, 31(7): 612-624.

冯斐, 吴小俊, 徐天阳. 基于子空间和直方图的多记忆自适应相关滤波目标跟踪算法 [J]. *模式识别与人工智能*, 2018, 31(7): 612-624.

[4] Bolme D, 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.

- [5] 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. Berlin, Heidelberg: Springer, 2012, 7575: 702-715.
- [6] 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.
- [7] Danelljan M, Häger G, Shahbaz K F, *et al.* Accurate scale estimation for robust visual tracking[C] // Proceedings of the British Machine Vision Conference 2014, September 1-5, 2014, Nottingham, UK. UK: BMVA Press, 2014.
- [8] He X D, Zhou S Z. Fast scale adaptive kernel correlation filtering algorithm for target tracking[J]. Laser & Optoelectronics Progress, 2018, 55(12): 121501.
何雪东, 周盛宗. 快速尺度自适应核相关滤波目标跟踪算法[J]. 激光与光电子学进展, 2018, 55(12): 121501.
- [9] 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.
- [10] Lukezic A, Vojir T, Zajc L C, *et al.* Discriminative correlation filter with channel and spatial reliability[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 4847-4856.
- [11] Lin Z C, Liu R S, Su Z X. Linearized alternating direction method with adaptive penalty for low-rank representation[C] // NIPS 2011: Neural Information Processing Systems Conference, December 12-15, 2011, Granada, Spain. Canada: NIPS, 2011: 612-620.
- [12] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [13] 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.
- [14] Lin B, Li Y. High-speed correlation filter tracking algorithm based on high-confidence updating strategy [J]. Acta Optica Sinica, 2019, 39(4): 0415003.
林彬, 李映. 基于高置信度更新策略的高速相关滤波跟踪算法[J]. 光学学报, 2019, 39(4): 0415003.
- [15] Wu Y, Lim J, Yang M H. Object tracking benchmark [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [16] Kristan M, Leonardis A, Matas J, *et al.* The visual object tracking VOT2016 challenge results[M] // Hua G, Jégou H. Computer vision workshops-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9914: 777-823.
- [17] 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, 2017: 6931-6939.
- [18] 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.
- [19] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking[C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 1144-1152.
- [20] Mao Z C, Chen H D. Long-term object tracking algorithm based on kernelized correlation filter[J]. Laser & Optoelectronics Progress, 2019, 56(1): 010702.
茅正冲, 陈海东. 基于核相关滤波的长期目标跟踪算法[J]. 激光与光电子学进展, 2019, 56(1): 010702.