

doi:10.3788/gzxb20184709.0910001

融合峰旁比和帧差均值自适应模型更新的视觉跟踪

熊昌镇¹, 车满强¹, 王润玲², 卢颜¹

(1 北方工业大学 城市道路交通智能控制技术北京市重点实验室, 北京 100144)

(2 北方工业大学 理学院, 北京 100144)

摘 要:为了让相关滤波模型更加适应目标外观的变化,提高相关滤波跟踪算法的鲁棒性和实时性,根据相关滤波响应值、帧差均值和目标运动位移之间的关系,提出了一种单层卷积相关滤波实时跟踪模型的自适应学习率调整跟踪方法.该方法首先选取单个卷积层卷积特征,减少了卷积特征维度,然后使用单层卷积特征训练相关滤波分类器预测目标位置,用快速尺度预测方法估计跟踪目标的尺度,并采用稀疏的模型更新策略,提高跟踪的速度;最后利用相关滤波预测响应图的峰旁比估计预测位置的可信度,结合图像帧差均值和目标的运动位移量来评估目标的表观变化,并根据目标预测的可信度和表观变化情况自适应调整相关滤波模型更新的学习率,使模型快速学习目标的变化特征,提高了目标跟踪的精度.在 OTB100 数据集上对算法进行测试,实验结果表明,本文算法的平均距离精度达 90.1%,优于实验中对比的 9 种主流算法,平均成功率值为 79.2%,仅次于 9 种算法中的连续卷积跟踪算法,平均速度为 31.8 帧/秒,是连续卷积相关滤波算法的近 30 倍.

关键词:目标跟踪;相关滤波;卷积特征;模型更新;峰旁比;帧差均值;学习率

中图分类号:TP491.4

文献标识码:A

文章编号:1004-4213(2018)09-0910001-12

Adaptive Model Update via Fusing Peak-to-sidelobe Ratio and Mean Frame Difference for Visual Tracking

XIONG Chang-zhen¹, CHE Man-qiang¹, WANG Run-ling², LU Yan¹

(1 Beijing Key Laboratory of Urban Intelligent Control, North China University of Technology, Beijing 100144, China)

(2 College of Sciences, North China University of Technology, Beijing 100144, China)

Abstract: In order to adapt the correlation filter model to the change of the target appearance and improve the robustness and real-time performance of the correlation filter algorithm for visual tracking, an adaptive learning rate adjustment method for real-time tracking of a single-layer convolution filter is proposed, which is based on the relationship of the correlation filter response value, mean frame difference and the object movement displacement. This method selects the convolution features of a single convolution layer to train the correlation filter classifier that is used to predict the position of object, reducing the convolution feature dimension and improving the speed of visual tracking. Meanwhile it uses the fast-scale prediction method to estimate the object's scale, and adopts a sparse model update strategy. Besides, the Peak-to-Sidelobe Ratio (PSR) of convolutional response is used to estimate the credibility of the predicted location. The apparent change of the object is evaluated by combining the mean frame difference and the object movement displacement. And the learning rate of the correlation filter model update can be adjusted by these two terms adaptively, so that the change characteristics of the object can be quickly learned. The accuracy of visual tracking is improved by this method. The method is tested on

基金项目:国家重点研发计划(No.2017YFC0821102)资助

第一作者:熊昌镇(1979-),男,副教授,博士,主要研究方向为深度学习、视频图像处理. Email:xczking@163.com

收稿日期:2018-05-06;录用日期:2018-06-12

<http://www.photon.ac.cn>

the standard OTB-100 dataset. The results show that the average distance accuracy is 90.1%, which is better than the state-of-the-art algorithms in the experiment. And the average success rate is 79.2%, which is only smaller than the continuous convolution tracking algorithm(CCOT). But the average speed is 31.8 frames per second, nearly 30 times of the CCOT.

Key words: Object tracking; Correlation filtering; Convolution feature; Model update; Peak-to-sidelobe ratio; Mean frame difference; Learning rate

OCIS Codes: 100.4996; 100.2000; 100.2960; 150.1135

0 引言

视觉跟踪是机器视觉、图像处理的重要分支,在医学、军事、智能交通、体育等方面有着广泛的应用.视觉跟踪的主要任务为:在视频序列的第一帧指定跟踪的目标,然后在后续帧连续估计目标的位置和尺度,其主要难点为目标运动过程中不定向和不同程度的运动会使目标发生旋转、尺度变化,以及遇到遮挡、背景变化、光照变化等一系列的干扰.因此,如何设计出鲁棒的实时目标跟踪算法来解决各种干扰下稳定跟踪目标的问题仍然面临着巨大的挑战.

自从 BOLME 等^[1]提出自适应相关滤波跟踪算法后,因该算法的高效实时性,受到了研究者的青睐,涌现出多种改进的相关滤波跟踪算法.为了提升相关滤波跟踪算法的精度,研究者们引入颜色属性(Color Name, CN)特征^[1]、方向梯度直方图(Histogram of Oriented Gradient, HOG)特征^[3-4]、融合 HOG 和 CN 特征^[5-7]、融合 HOG 和颜色直方图特征^[8](Sum of Template And Pixel-wise Learners, Staple)等人工设计的特征及组合特征进行跟踪;LUKEZIC 等^[9]分别针对图像前景和背景提取颜色直方图特征,然后对提取到的特征进行空域的可靠性判断,降低背景样本造成的影响;DANELLJAN 等^[4]在损失函数中引入空间正则化惩罚项(Spatially Regularized Correlation Filter, SRDCF),减少由循环矩阵引起的边缘效应.DANELLJAN 等^[10]提出了快速尺度估计的跟踪算法(Discriminative Scale Space Tracking, DSST)来提高目标定位精度.HAMED 等^[11]利用背景信息对相关滤波跟踪算法进行改进,在损失部分跟踪速度的情况下提升了跟踪的精度.

随着深度学习的快速发展,卷积神经网络提取的包含丰富语义和空间信息的特征也应用于相关滤波跟踪领域来提高跟踪的精度.MA 等^[12]提出分层卷积跟踪算法(Hierarchical Convolutional Features for visual Tracking, HCFT),利用预训练好的网络模型提取不同深度的卷积特征,融合分层卷积特征相关滤波预测目标位置;后来作者对该算法进行改进(HCFTS)^[13],融入尺度估计,并利用 HOG 特征在目标跟踪失败时进行重定位,使算法在跟踪精度方面得到很好的提升,但速度达不到实时;QI 等^[14]提出多层卷积相关滤波跟踪算法(Hedged Deep Tracking, HDT),在相关滤波跟踪框架下使用多层卷积特征进行跟踪;DANELLJAN 等^[5]提出连续卷积跟踪算法(Continuous Convolution Operators for visual Tracking, CCOT),融合卷积特征、HOG 特征和 CN 特征训练多个滤波器进行跟踪,跟踪精度高,但跟踪速度低于每秒 1 帧.SUN 等^[14]提出一种基于联合判别以及可靠性学习的相关滤波跟踪算法,将目标图像进行分块提取特征,然后自适应学习权重进行融合预测目标位置,跟踪精度极高,但跟踪速度慢.ZHU 等^[16]在卷积特征中融入包含时空信息的光流特征进行跟踪,跟踪性能好,但也达不到实时.

深度特征的通道数多,容易造成跟踪速度缓慢,为了提高跟踪的速度,DANELLJAN 等^[17]在 CCOT 算法的基础上采用因式分解对特征进行降维,融合稀疏的卷积特征、HOG 特征和 CN 特征进行跟踪,同时改变样本的分布形式,使跟踪精度和速度都得到提升,但还达不到实时;CAI 等^[18]利用主成分分析(Principal Component Analysis, PCA)对深度卷积特征进行降维,然后在 SRDCF 的跟踪框架下利用稀疏卷积特征进行跟踪,有效提升了跟踪性能,但跟踪精度有待进一步提高;WANG 等^[19]在分层卷积相关滤波跟踪算法的基础上提出了多尺度域跟踪算法(Multi-Scale Domain Adaptation, MSDAT),将卷积特征的维度降为原来的 1/8,同时采用稀疏模型更新策略,提升了跟踪的实时性,但跟踪精度有所下降;FAN 等^[20]提出一种并行验证跟踪算法,使用相关滤波模型进行跟踪,同时利用全卷积孪生网络跟踪算法在跟踪失败时进行重新定位,提升了算法性能,但重定位算法比较耗时.

除了各种特征的组合使用和相关滤波自身方面的改进外,也有不少算法通过估计预测目标位置的可靠性来选择性更新模型,从而提升算法的鲁棒性,GALOOGAHI 等^[21]提出有限边界的相关滤波来降低边界效

应,并采用峰旁比(Peak-to-Sidelobe, PSR)估计相关滤波响应预测的可靠性,改进模型的更新策略;YANG等^[22]利用最佳伙伴相似性改进空间正则判别相关滤波跟踪算法,通过峰旁比判断目标的置信度,解决遮挡情况下的模板更新问题,并用模板匹配重定位算法提高跟踪精度;WANG等^[23]提出基于结构化支持向量机分类器使用深度特征的目标跟踪算法(Large Margin Object Tracking, D_LMCF),通过多峰检测来解决相似物体的干扰,通过平均峰值相关能量(Average Peak-to-Correlation Energy, APCE)判断当前帧是否出现了遮挡、以及目标丢失,选择性更新模型,提升了算法的鲁棒性;ZHU等^[24]将特征提取和跟踪过程作为卷积操作联合训练,使学习到的卷积特征与跟踪过程紧密结合,减小了卷积特征提取与分类器训练的耗时,并引入峰值噪声比(Peak Noise Ratio, PNR)对模型进行更新,稀疏了模型更新次数,使得算法达到实时跟踪,但跟踪的精度还有提升的空间。

由于多种特征的组合使用和相关滤波自身模型的改进,使得跟踪的准确度得到了很好的提升,尽管目标预测可靠性估计和稀疏更新策略使深度卷积特征的跟踪速度有所提升,但很少有算法能达到实时跟踪效果。因此,本文从降低特征维度、模型更新策略和学习率调整三方面来提升跟踪算法的性能,主要有以下工作:1)选取能表征目标的单层卷积特征,降低模型的复杂度,提升算法的实时性;2)采用稀疏模型更新策略来防止模型漂移,提升目标跟踪的实时性;3)使用预测响应图的峰旁比估计预测位置的可信度,并用图像帧差均值与目标的位移量来评估目标的表现变化,结合预测可信度和表现变化综合调整模型的学习率,使模型快速适应目标的表现变化,解决卷积特征层数减小造成的影响,提高算法的精度。

1 单层卷积相关滤波跟踪算法

1.1 卷积特征选择

卷积特征与传统手工特征最大的区别为:卷积神经网络模型通过大量的标注数据训练得到,在不同数据集上训练得到的模型具有很好的泛化和迁移能力,利用训练好的网络模型提取目标图片的卷积特征时,得到的特征包含丰富的语义信息,具有很强的判别性。VGG-19网络模型是在ImageNet数据集上训练的分类模型,可以输出5组卷积层的卷积特征,这5组卷积特征包含不同目标信息,低层特征保存了更多的细粒度的空间特征,随着网络深度的增加,特征图分辨率不断降低,但是包含更丰富的语义信息,更容易区分类别之间的差异。在目标跟踪中,希望所提取的特征既能够有很好的语义特征,以便于更好地适应目标的表现变化,又能保留准确的空间信息,以便于对目标进行定位,防止模型漂移现象的发生,因此选用中高层的单层卷积特征用于跟踪,减少特征层数。图1为VGG-19的第4组卷积层的第4个卷积操作(c4-4)、第4个pool层(p4)和第5组卷积层的第1个卷积操作(c5-1)得到的卷积特征加余弦窗后的可视化结果,可以看出这些中高层特征既保留一部分低层信息也包含高层特征的语义信息,都可很好地定位到目标,但这些卷积层的特征对不同视频的反应有所不同,因此本文通过实验验证选取p4单层的卷积特征用于目标跟踪。

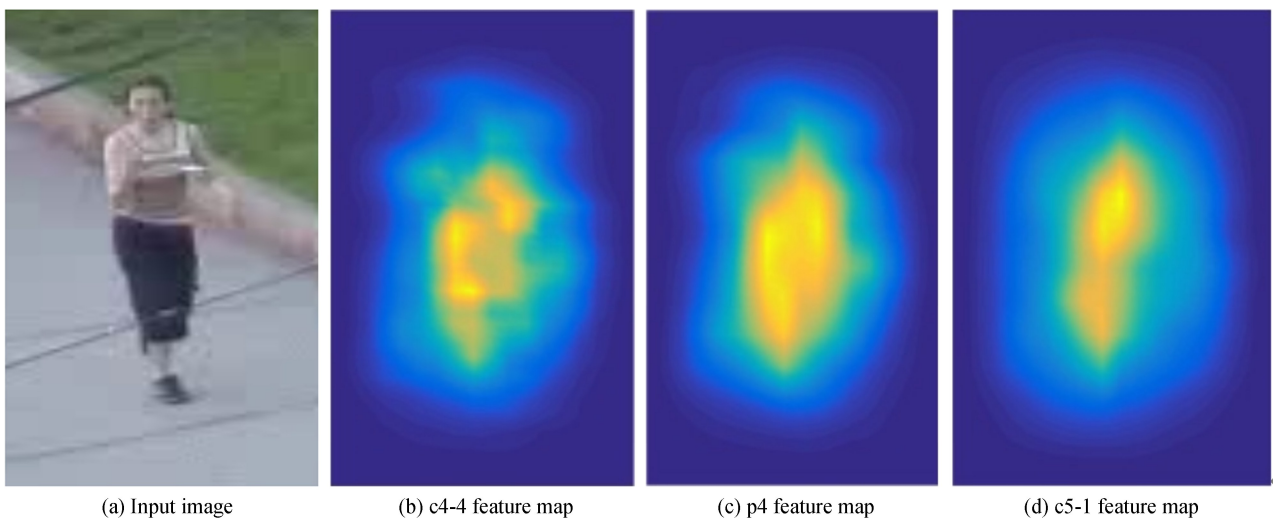


图1 VGG-19中不同卷积层特征图的可视化结果

Fig.1 Visualization for different convolutional layers of VGG-19

1.2 相关滤波跟踪算法

相关滤波跟踪算法主要包括分类器训练、目标预测和模型更新三部分内容.训练分类器时采用二维高斯分布的类标函数 y

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

式中, $y(m, n)$ 为 (m, n) 处像素的标签函数, M, N 代表目标块的宽和高, $\sigma > 0$ 为高斯核带宽.对提取的第一帧目标区域的卷积特征 $\mathbf{x}_d (d=1, 2, \dots, D)$, 使用循环平移的稠密采样技术和傅里叶变换域快速求解最小化目标函数

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

得到分类器 \mathbf{W}_d

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

式中, $x_{m, n}$ 为图像在 (m, n) 处的特征值, \cdot 表示内积, \mathbf{Y} 为标签函数 y 的傅里叶变换, $\overline{\mathbf{X}}$ 为 \mathbf{X} 复共轭, \odot 表示对应元素相乘, D 代表特征的通道数; $\lambda > 0$ 为优化求解时正则项参数.

在新一帧的图片时,提取新输入图像块 pool4 层的卷积特征 \mathbf{z}_d , 计算傅里叶变换后的卷积特征 \mathbf{Z}_d 与分类器 \mathbf{W}_d 的相关响应输出 \mathbf{f}

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

式中, F^{-1} 为傅里叶逆变换, 寻找响应图 \mathbf{f} 最大值的位置即为预测的目标中心位置, 并使用快速尺度估计算法预测目标的尺度^[9].

为使模型适应目标运动过程中表观模型的变化, 利用新检测到的图像块特征更新分类器 \mathbf{W}_d . 为提高算法的跟踪速度, 避免模型过拟合造成的目标漂移, 受文献[16]启发, 采用每两帧更新一次模型的策略. 具体更新方式为分别更新相关滤波分类器的分子 \mathbf{A}_d 和分母 \mathbf{B}_d , 即

$$\mathbf{W}_{d, t} = \frac{\mathbf{A}_{d, t}}{\mathbf{B}_{d, t} + \lambda} \quad (5)$$

$$\mathbf{A}_{d, t} = (1 - \eta) \mathbf{A}_{d, t-2} + \eta \mathbf{Y} \odot \overline{\mathbf{X}}_{d, t} \quad (6)$$

$$\mathbf{B}_{d, t} = (1 - \eta) \mathbf{B}_{d, t-2} + \eta \sum_{i=1}^D \mathbf{X}_{d, t}^i \odot \overline{\mathbf{X}}_{d, t}^i \quad (7)$$

式中, t 为帧序列, η 为学习率. 为进一步提高模型对目标表观模型变化的适应, 提高算法的性能, 在稀疏更新的基础上, 融合峰旁比、帧差均值和目标运动位移综合调整模型的学习率 η .

2 自适应分段学习速率调整算法

单层卷积相关滤波跟踪模型属于判别式在线跟踪模型, 首先利用第一帧图片训练相关滤波器分类器, 然后在后续帧中根据预测的结果对模型进行适时、适度的更新, 具有较强的灵活性, 可以不断地融入目标和背景的变化信息, 使得模型适合当前目标的形态. 模型更新过程主要通过文中式(5)、(6)、(7)完成, 其中学习速率 η 表示模型的学习能力. 尽管有一些算法对预测目标的可靠性进行估计, 对于目标预测可靠性高的视频帧, 进行模型更新, 对目标预测可靠性低的视频帧不进行模型更新^[14, 16-17], 避免了部分目标模型漂移的现象, 但现有的算法通常采用固定的学习速率调整模型的变化, 无法适应所用视频的变化情况.

实际视频跟踪过程中, 目标随时都有可能发生不同程度、不定向的运动, 可能出现各种各样的干扰, 当目标自身发生较大改变, 例如非刚性形变、平面内旋转、平面外旋转、尺度变化等时, 模型需要做出较大调整来适合当前目标的形态, 此时需要较大的学习速率; 当目标自身变化较小, 例如由周围环境引起的光照变化、缓慢的背景运动、遮挡、超出视野等时, 此时学习速率过快, 容易导致模型的漂移, 模型不宜做出较大调整, 需要较小的学习速率.

视觉跟踪算法中除了给定的第一帧外, 其它帧的位置都是预测出来, 为了调整模型的更新速率, 就需要

判断预测位置的可靠性,以及目标形态的变化程度.可选用 PSR 来评估预测位置的可靠性;对摄像机抖动、运动模糊、光照条件变化、大尺度运动等造成的表观模型变化,可使用相邻帧图像的差值来评估,当帧差差异变化大时说明目标的表观模型发生了大的变化,如大尺度的运动或瞬间光照变化等,当帧差差异小时,说明目标运动缓慢或是发生遮挡,目标的表观模型变化不大.因此本文融合相邻两帧图像帧差均值、预测响应图的峰旁比、相邻两帧预测位置的位移量来评估模型的可信度和表观变化,综合调整目标跟踪模型的学习率.

具体算法步骤为:

Step1 计算当前帧预测的相关滤波响应图的峰旁比.对于第 t 帧分类器预测的响应 f_t ,其峰旁比为

$$P_{sr}^t = \frac{\max(f_t) - \mu_t}{\sigma_t} \quad (8)$$

式中, μ_t 和 σ_t 分别为响应图 f_t 的均值和方差.

Step2 计算相邻 2 帧图像的帧差均值.对于输入大小为 $M \times N$ 的第 t 帧图像,可计算其与前 1 帧图像帧差均值

$$e_t = \frac{\sum_{m,n}^{M,N} |J_{m,n}^t - J_{m,n}^{t-1}|}{MN} \quad (9)$$

式中, $J_{m,n}$ 表示图像 J 在 (m,n) 处的像素值.

Step3 计算相邻 2 帧预测位置的距离.得到第 t 帧预测的目标位置 (x_t, y_t) ,可得到当前预测目标位置与上帧目标位置之间的运动位移量为

$$s_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2} \quad (10)$$

式中, (x, y) 为目标中心位置的坐标.

Step4 分段学习速率调整.峰旁比可以作为模型预测定位的准确性的判据,当峰旁比 $P_{sr}^t < \delta$ 时,表示模型未能定位到目标,不需要更新模型,学习速率为 $\eta = \eta_1$.

当 $P_{sr}^t \geq \delta$ 时,表示模型定位到目标,需根据帧差均值判断目标表观模型的变化来调节学习率.

1) 当 $e_t < \gamma_1$,可认为相邻两帧图像之间变化较小,目标的表观模型变化也较小,使用较低的学习率 $\eta = \eta_2$;

2) 当 $\gamma_1 \leq e_t < \gamma_2$ 时,可认为相邻两帧图像之间变化适中,有可能是背景发生改变,或是目标自身发生改变,两种情况都需要适中的学习率更新模型,即 $\eta = \eta_3$;

3) 当 $e_t \geq \gamma_2$ 时,表明相邻两帧图像之间变化大,有可能是目标背景区域变化大,或是目标自身变化大,因此通过预测目标位置的位移量来判断是否发生了大的位移.

a) 当 $s_t < \zeta$ 时可认为是背景发生变化,选用适中的学习率 $\eta = \eta_3$;

b) 当 $s_t \geq \zeta$ 可认为是目标发生变化,此时需要较大的学习率 $\eta = \eta_4$ 调整模型来适合当前目标的表现.

具体设置的分段学习速率为

$$\eta = \begin{cases} \eta_1 & P_{sr}^t < \delta \\ \eta_2 & P_{sr}^t \geq \delta, e_t < \gamma_1 \\ \eta_3 & P_{sr}^t \geq \delta, \gamma_1 < e_t < \gamma_2 \\ \eta_3 & P_{sr}^t \geq \delta, e_t \geq \gamma_2, s_t < \zeta \\ \eta_4 & P_{sr}^t \geq \delta, e_t \geq \gamma_2, s_t \geq \zeta \end{cases} \quad (11)$$

3 实验结果

3.1 实验环境与参数设置

为验证本文算法的有效性,选取 OTB-100 标准数据集对算法进行评估,该数据集总共包含 100 组不同类型的视频序列,视频中包含遮挡、快速运动、模糊等 11 种干扰情况.实验中选用中心位置误差 (Center Location Error, CLE)、距离精度 (Distance Precision, DP)、重叠精度 (Overlap Precision, OP)、一次性通过 (One-Pass Evaluation, OPE) 的距离精度曲线 (precision plot) 和成功率曲线 (success plot) 对算法进行评估.

其中, CLE 指检测到的目标中心位置与目标真实中心位置的平均欧式距离; DP 指中心位置误差小于 20 像素的帧数占视频总帧数的百分比; OP 指边界框重叠的位置超过 0.5 的帧数占视频总帧数的百分比; 距离精度曲线指显示距离精度值与阈值之间关系的曲线图, 成功率图指显示边界框重叠率与阈值之间的关系图. 实验平台为 ubuntu16.04 系统下的 MATLAB R2015b, 所有的实验均在配置为 Intel Core i7-7800XCPU, GTX1080Ti GPU, 内存为 16GB 的台式电脑上完成. 本文算法在 OTB-100 标准数据集上测试的具体参数为: 选用的卷积层为 VGG-19 的 pool4, 相关滤波分类器所用的正则化参数为 10^{-4} , 类标签函数的高斯核带宽 $\sigma=0.225$, 模型更新策略为每两帧更新一次, 定位准确性判断阈值 $\delta=4.13$, 帧差阈值 $\gamma_1=2.5, \gamma_2=9.65$, 位移量判断阈值 $\zeta=11$, 分段学习速率为 $\eta_1=0, \eta_2=0.05, \eta_3=0.01, \eta_4=0.1$, 快速尺度估计相关滤波器相关参数与原文的参数一致^[9].

3.2 自身算法的对比结果

本文主要通过以下 3 点提升相关滤波跟踪算法的性能: 1) 在相关滤波跟踪框架下使用单层卷积特征进行跟踪; 2) 更新相关滤波模型时选用每 2 帧更新一次的策略; 3) 提出新颖的自适应学习率调整算法. 为了验证本文算法中各策略的效果, 分析选用不同卷积层、稀疏更新和不同学习率调整对目标跟踪算法的影响, 显示了实验过程中的部分跟踪数据. 表 1 显示了 VGG19 第 4 组卷积层的第 4 个卷积操作的特征 c4-4、和第 4 个 Pooling 层的卷积特征 p4、第 5 组卷积层第一个卷积操作的特征 c5-1 的单个卷积层在固定学习率 0.01 每帧都进行模型更新(c4-4, p4, p5-1)和每 2 帧更新一次模型下的结果(c4-4_s, p4_s, c5-1_s), 以及单独使用 p4 层在自适应学习率调整和稀疏模型更新下的结果(ours). 从表中的数据可以看出, 稀疏模型更新策略在使用 c4-4、p4 和 c5-1 层的跟踪精度和跟踪速度都有所提升, 其中 p4 层的提升效果最明显, 因此在稀疏模型更新选用单层的 p4 层卷积特征的跟踪基础上, 加入自适应学习率调整算法后, 可以看出在速度损失仅 2.6/3 的情况下, 平均 DP 达到 90.1%, 说明自适应学习率调整策略很好地提升了单层相关滤波跟踪算法的鲁棒性.

表 1 不同卷积特征选择及策略选用跟踪结果

Table 1 Tracking results of different convolution feature and strategy selection

	c4-4	p4	c5-1	c4-4_s	p4_s	c5-1_s	Ours
Average DP/%	83.2	83.7	85.6	85.5	87.7	86.4	90.1
Speed/fps	25.2	26.2	26.2	32.6	33.4	33.7	31.8

注: 文中表中排名第 1 的数据均用粗体实下划线表示, 排名第 2 的数据均用粗体虚下划线表示.

图 2 为 p4 卷积层在不同学习率调整策略组合下的 OPE 准确度和成功率图, 其中 p4_s 为 p4 单层每两帧更新一次模型时的跟踪结果, psr 为仅使用峰旁比调整学习率时的跟踪结果, 即

$$\eta = \begin{cases} 0 & P'_{sr} < 4.13 \\ 0.01 & P'_{sr} \geq 4.13 \end{cases} \quad (12)$$

psr+md 为峰旁比与相邻两帧图像差均值调整学习率时的跟踪结果, 即

$$\eta = \begin{cases} 0 & P'_{sr} < 4.13 \\ 0.005 & P'_{sr} \geq 4.13, e_t < 2.5 \\ 0.01 & P'_{sr} \geq 4.13, 2.5 \leq e_t \leq 9.65 \\ 0.1 & P'_{sr} \geq 4.13, e_t \geq 9.65 \end{cases} \quad (13)$$

Ours 为使用式(11)综合调节学习率的跟踪结果, 可以看出在随着策略的结合使用准确度和成功率都呈上升趋势, 当使用文中(11)式综合调整学习率时跟踪精度最高, 图 2 中也显示了对应策略时的跟踪速度, 可以看出学习率调整算法几乎不影响跟踪的速度, 仍然可达到实时的效果. 图 3 显示了选用 p4 层卷积特征在不同学习率调整算法下 4 组视频的可视化结果, 可以看出使用式(11)进行自适应学习率调整的算法都可稳定的跟踪到目标.

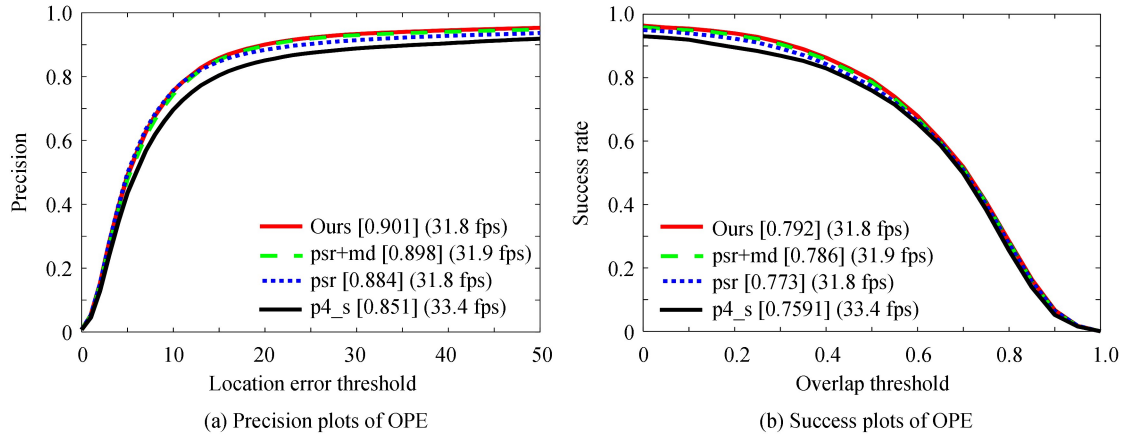


图2 不同学习率调整策略组合的 OPE 准确度和成功率

Fig.2 Precision plots and success plots of OPE for combination of different learning rate adjustment method

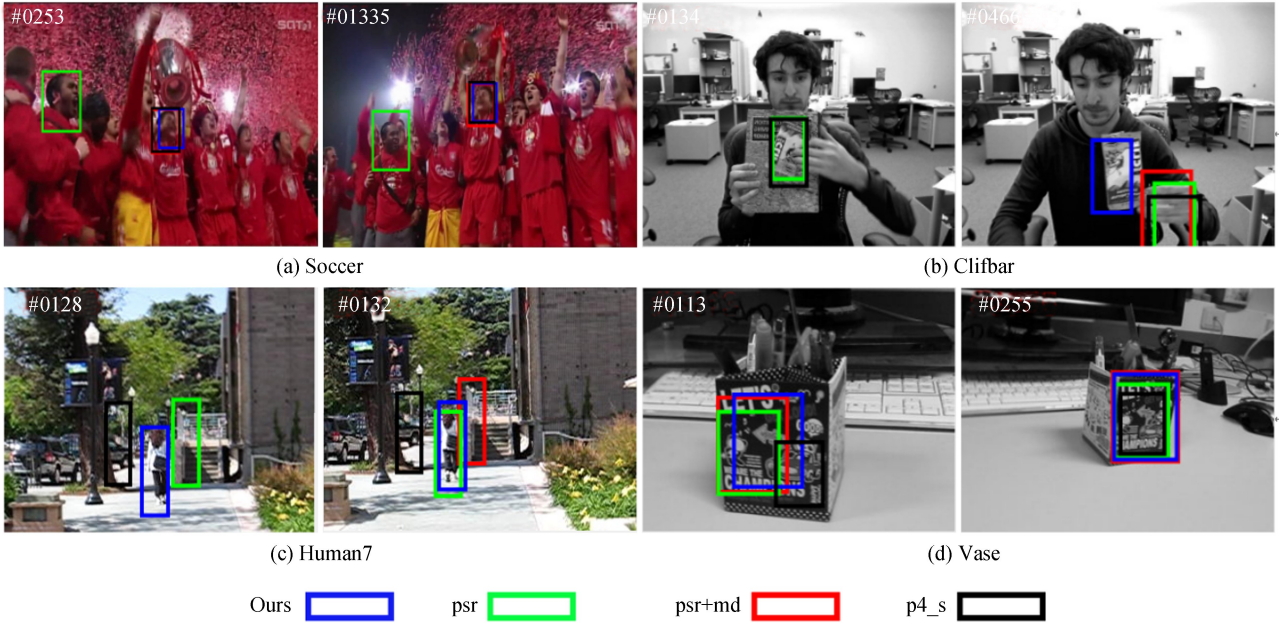


图3 不同学习率调整策略组合的部分跟踪结果对比

Fig.3 Partial tracking results for combination of different learning rate adjustment method

以上实验表明,单层卷积特征、稀疏更新策略和自适应学习率速率调整的综合应用使算法跟踪精度和实时性得到了很好的平衡。

3.3 与现有主流算法对比结果

为了进一步验证算法的整体性能,选取现有的 9 种主流跟踪算法进行对比,分别是 CCOT^[5]、HCFT^[12]、HCFTS^[13]、HDT^[14]、MSDAT^[19]、Staple^[8]、SRDCF^[4]、DSST^[10]、D_LMCF^[23],其中 CCOT、HCFT、HCFTS、HDT、MSDAT、D_LMCF 使用了深度卷积特征,CCOT、DSST、Staple、HCFTS4 几种算法带尺度估计.在 OTB-100 数据集上对比的 OPE 精确度和成功率曲线如图 4 所示,OPE 距离精度方面,比 HCFT 算法高 6.7 个百分点,比 HCFT 的改进算法 HCFTS 高 3.1 个百分点,且原论文 HCFTS 的跟踪速度仅为 6.7 frame/s,本文算法的跟踪速度为 31 frame/s,远高于 HCFTS 算法.OPE 成功率方面,本文算法仅次于 CCOT 算法,但 CCOT 算法原文速度仅 1 frame/s 左右,本文算法速度远高于该算法。

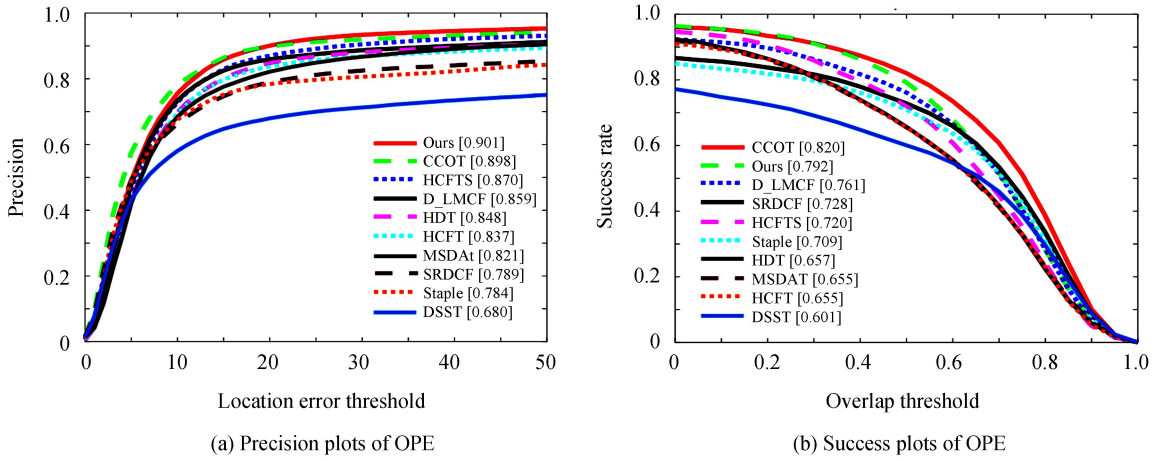
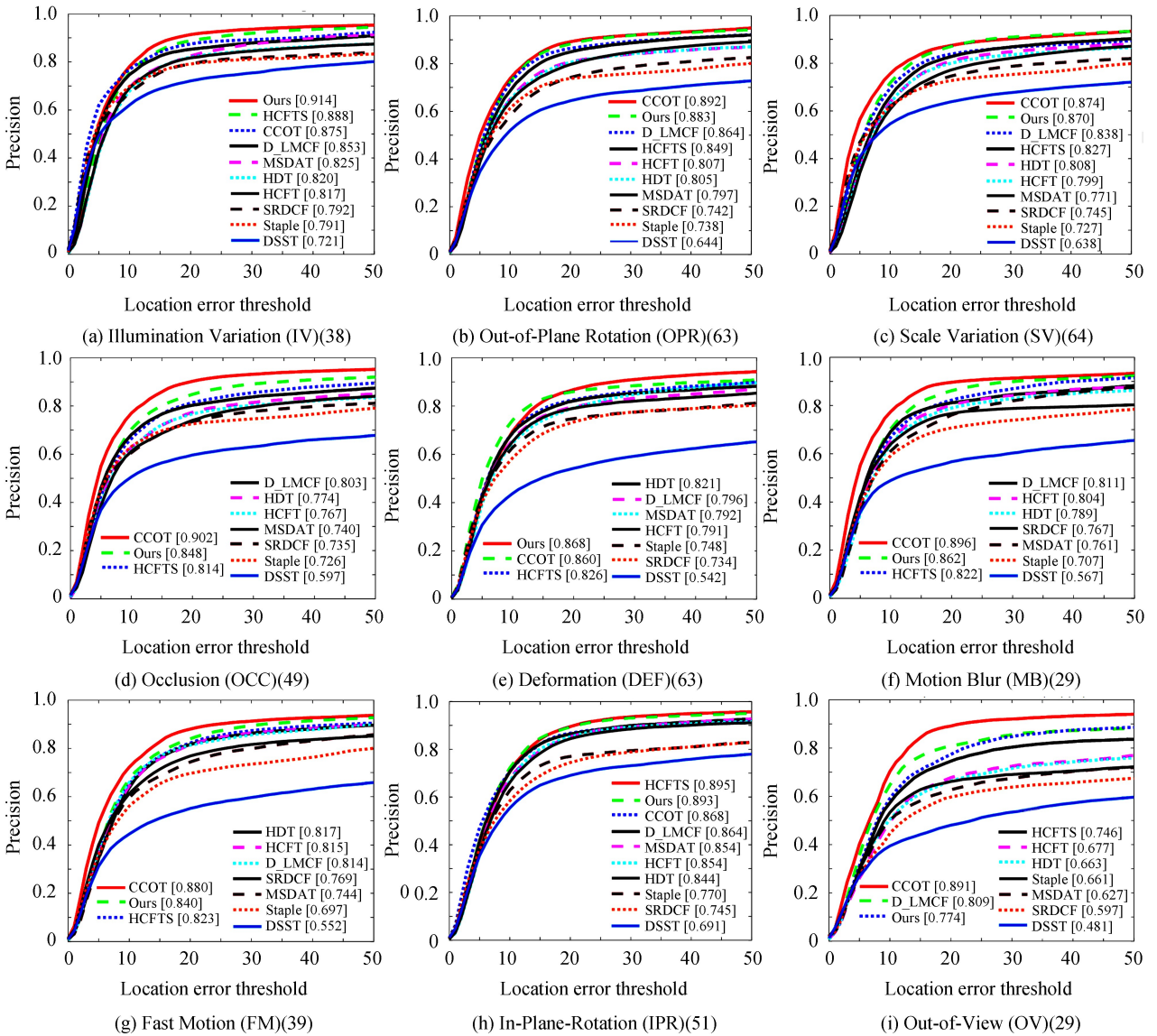


图 4 10 种算法的距离精度曲线和成功率曲线

Fig.4 Precision and success plots of ten algorithms on OTB-100

图 5 和图 6 为 OTB-100 数据集给出的不同挑战下的跟踪结果,图 5 精确度对比中,本文算法在 BC、DEF、IV 三种挑战下位居第一,在除 OV 的 7 种挑战下本文算法位居第二;图 6 成功率对比中,BC、OPR、



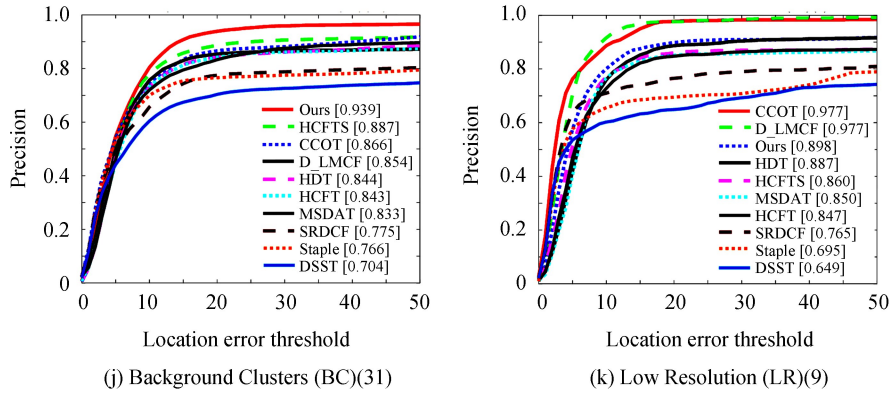
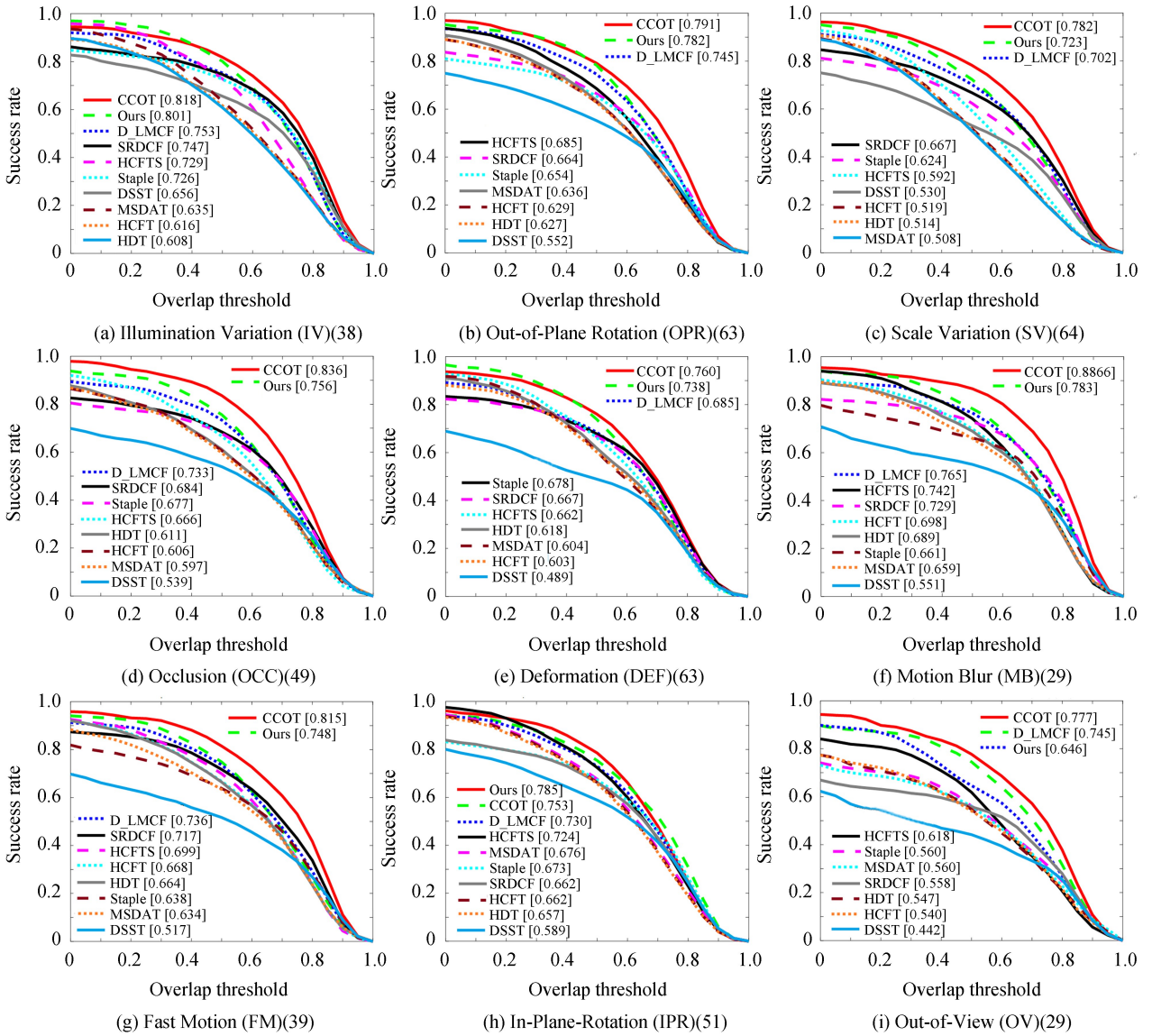


图 5 不同算法在 11 种属性上的准确度图(标题上数字表示此种属性的数据集数量)
 Fig.5 Precision plots of videos with eleven different attributes for different trackers
 (The number in the title indicates the number of sequences)



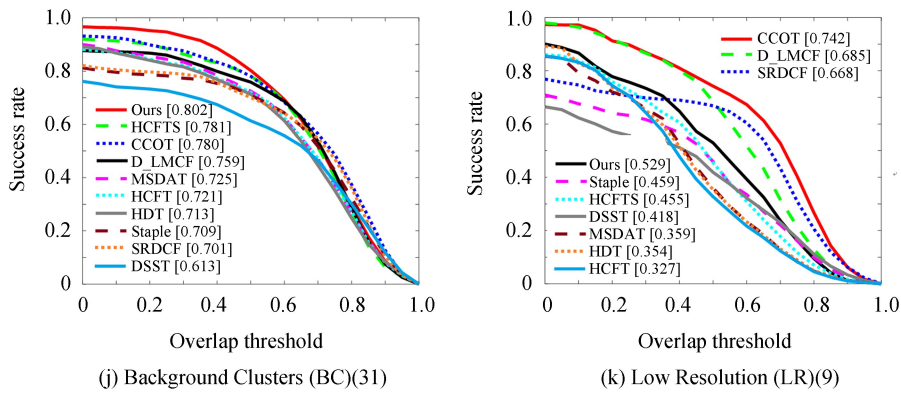


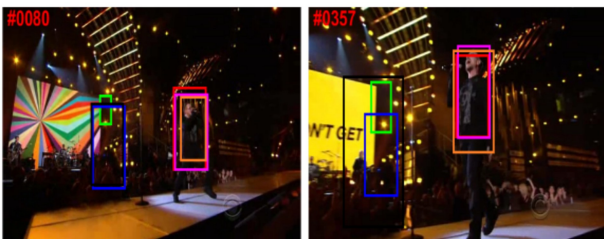
图 6 不同算法在 11 种属性上的准成功率图(标题上数字表示此种属性的数据集数量)
 Fig.6 Success plots of videos with eleven different attributes for different trackers
 (The number in the title indicates the number of sequences)

OPR 三种挑战下本文算法位居第一,在除 LR 的其余 7 中挑战下位居第二.说明本文自适应学习率的调整很好地提升了算法对于目标变化的学习能力,以及在背景复杂时模型的适应能力,其次在精确度和成功率对比中,本文在 7 种挑战中位居第二,进一步说明了本文算法整体性能的稳定性,其中这 7 种挑战中主要为 CCOT 算法位居第一,但本文算法速度远高于 CCOT 算法.可以看出本文算法的整体性能优于对比的 9 种算法.

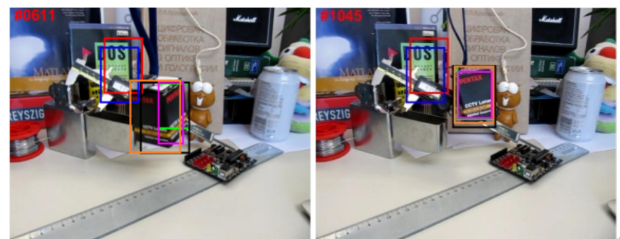
为了更好地说明本文算法在复杂环境下的跟踪性能,选取典型视频的部分帧跟踪结果定性分析本文算法的跟踪性能,选取的视频为:singer2、box、human3、girl2、human2、vase 共 6 组视频,所包含的跟踪挑战如表 2 所示(表中的英文简写为图 5-6 中不同属性全称的简写),可以看出 6 组视频包含了 OTB 数据集中可能出现的所有挑战,从定量对比实验中可以看出本文算法性能远高于 DSST、Staple、SRDCF、HDT 四种算法,因此本实验中仅与剩余的 5 种算法进行比较,跟踪结果如图 7 所示.在图 7 中展示的 6 组视频的跟踪结果中,除本文跟踪算法外的其他 5 种算法均出现跟丢或者尺度预测不准确的现象,只有本文算法稳定地跟踪到展示的 6 组视频,可以看出当目标在运动过程中遇到形变、快速运动、尺度变化、背景混乱等干扰时,本文自适应学习率调整方法可以提升算法模型的适应能力,整体性能也优于对比的 5 种算法.

表 2 测试视频的属性
 Table 2 The properties of test videos

Video	Frames	Size of object	Challenging aspects
Singer2	366	122×67	IV, DEF, IPR, OPR, BC
Box	1 161	111×80	IV, SV, OCC, MB, IPR, OPR, OV, BC, LR
Human3	1 698	69×37	SV, OCC, DEF, OPR, BC
Firl2	1 500	171×44	SV, OCC, DEF, MB, OPR
Human2	1 128	162×47	IV, SV, MB, OPR
Vase	1 128	59×45	SV, FM, IPR



(a) Singer2



(b) Box

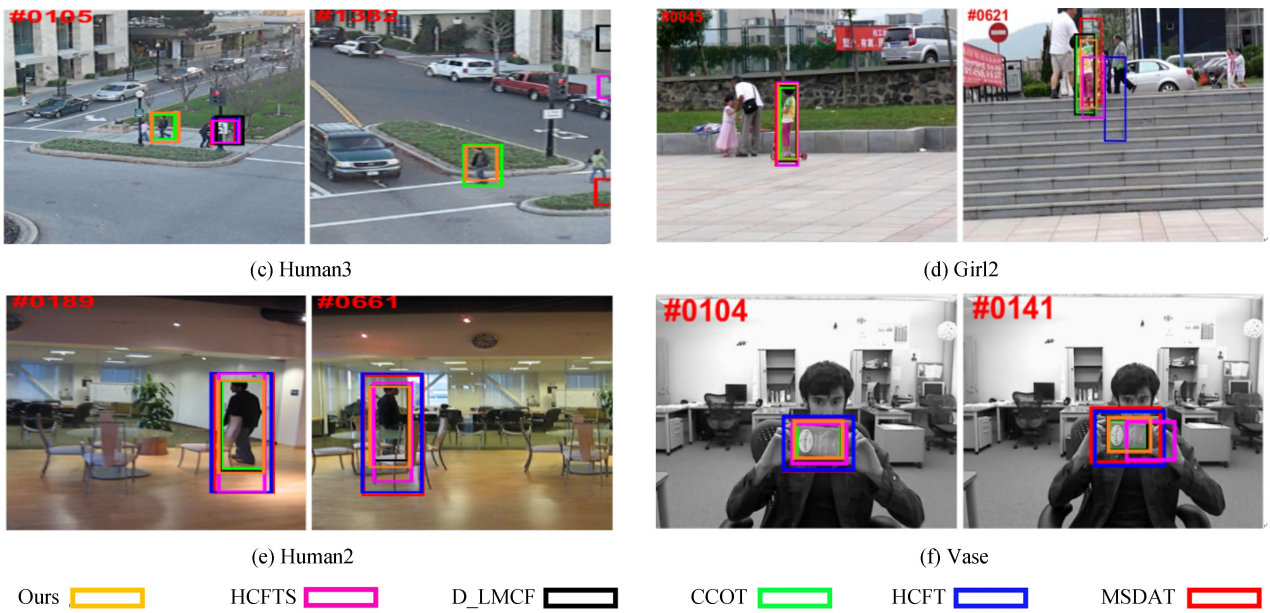


图7 6种跟踪算法的部分跟踪结果对比
Fig.7 Comparison of partial tracking results of six trackers

4 结论

在分层卷积相关滤波跟踪算法的基础上,对VGG-19的网络结构进行分析,将原三层卷积特征融合跟踪算法改进为单层卷积相关滤波跟踪算法,根据视频序列中图像变化的规律,将原算法每帧都进行模型更新的方式改为每两帧更新一次,同时结合峰旁比、帧差均值和位移量三个评估指标提出了一种自适应学习率调整算法,提升了相关滤波模型的学习能力和目标遇到形变时的定位能力.在OTB-100数据集上对算法进行测试,实验中分析不同卷积层、不同模型更新策略和综合学习率调整算法的跟踪效果,同时与其它9种主流的算法进行了对比实验,结果表明本文算法的平均距离精度达90.1%,平均速度为31.8帧/秒,整体性能优于对比的9种跟踪算法,验证了提出的自适应模型更新学习率调整算法的有效性,显示了本文算法在目标发生剧烈形变、背景混乱等情况时能稳定实时地跟踪到目标.

参考文献

- [1] 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, Piscataway: IEEE, 2010: 2544-2550.
- [2] DANELLJAN M, KHAN F S, FELSBERG M. Adaptive color attributes for real-time visual tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos: Washington, DC: IEEE Computer Society Press, 2014: 1090-1097.
- [3] HENRIQUES J F, CASEIRO R, MARTINS P, *et al.* High-speed tracking with kernelized correlation filters[J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2015, **37**(3): 583-596.
- [4] DANELLJAN M, HAGER G, KHAN F S, *et al.* Learning spatially regularized correlation filters for visual tracking[C]. IEEE International Conference on Computer Vision, Los Alamitos: Washington, DC: IEEE Computer Society Press, 2015: 4310-4318.
- [5] DANELLJAN M, ROBINSON A, KHAN F S, *et al.* Beyond correlation filters: learning continuous convolution operators for visual tracking[C]. European Conference on Computer Vision, Verlag: Berlin:Springer, 2016: 472-488.
- [6] DANELLJAN M, BHAT G, KHAN F S, *et al.* ECO: Efficient convolution operators for tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos: Washington, DC: IEEE Computer Society Press, 2017: 6931-6939.
- [7] XIONG Chang-zhen, ZHAO Lu-lu, GUO Fen-hong. Kernelized correlation filters tracking based on adaptive feature fusion[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2017, **29**(6): 1068-1074.
熊昌镇, 赵璐璐, 郭芬红. 自适应特征融合的核相关滤波跟踪算法[J]. 计算机辅助设计与图形学学报, 2017, **29**(6): 1068-1074.
- [8] BERTINETTO L, VALMADRE J, GOLODETZ S. Staple: complementary learners for real-time tracking[C]. IEEE

- International Conference on Computer Vision and Pattern Recognition, Los Alamitos: IEEE Computer Society Press, 2016: 1401-1409.
- [9] LUKEZIC A, VOJIR T, ZAJC L C, *et al.* Discriminative correlation filter with channel and spatial reliability[C]. IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2017: 4847-4856.
- [10] DANELLJAN M, HÄGER G, SHAHBAZ KHAN F, *et al.* Accurate scale estimation for robust visual tracking[C]. British Machine Vision Conference, 2014: 1-11.
- [11] HAMED KI G, ASHTON F, SIMON L. Learning background-aware correlation filters for visual tracking[C]. IEEE International Conference on Computer Vision (ICCV), Los Alamitos: Washington, DC: IEEE Computer Society Press, 2017: 1144-1152.
- [12] MA C, HUANG J B, YANG X, *et al.* Hierarchical convolutional features for visual tracking[C]. IEEE International Conference on Computer Vision (ICCV), Los Alamitos: Washington, DC: IEEE Computer Society Press, 2015: 3074-3082.
- [13] MA C, HUANG J B, YANG X, *et al.* Robust visual tracking via hierarchical convolutional features [OL]. [2018-4-20]. <https://arxiv.org/abs/1707.03816v1>.
- [14] YUANKAI Q I, SHENGPING ZHANG, LEI QIN, *et al.* Hedged deep tracking [C]. IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos: Washington, DC: IEEE Computer Society Press, 2016: 4303-4311.
- [15] SUN C, WANG D, LU H, *et al.* Correlation tracking via joint discrimination and reliability learning[OL]. [2018-6-8] <https://arxiv.org/pdf/1804.08965.pdf>.
- [16] ZHU Z, WU W, ZOU W, *et al.* End-to-end flow correlation tracking with spatial-temporal attention[OL]. [2018-6-8]. <https://arxiv.org/pdf/1711.01124.pdf>.
- [17] DANELLJAN M, BHAT G, KHAN F S, *et al.* ECO: Efficient convolution operators for tracking [C]. IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos: IEEE Computer Society Press, 2017: 6931-6939.
- [18] CAI Yu-zhu, YANG De-dong, MAO Ning, *et al.* Visual tracking algorithm based on adaptive convolution features [J]. *Acta Optica Sinica*, 2017, **37**(03): 0315002.
蔡玉柱, 杨德东, 毛宁, 等. 基于自适应卷积特征的目标跟踪算法[J]. 光学学报, 2017, **37**(03): 0315002.
- [19] WANG X, LI H, LI Y, *et al.* Robust and real-time deep tracking via multi-scale domain adaptation [C]. IEEE International Conference on Multimedia and Expo, Los Alamitos: Washington, DC: IEEE Computer Society Press, 2017: 1338-1343.
- [20] FAN H, LING H. Parallel tracking and verifying: a framework for real-time and high accuracy visual tracking [C]. IEEE International Conference on Computer Vision (ICCV), Los Alamitos: Washington, DC: IEEE Computer Society Press, 2017: 5487-5495.
- [21] GALOOGAHI H K, SIM T, LUCEY S. Correlation filters with limited boundaries[C]. IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2015: 4630-4638.
- [22] YANG De-dong, MAO Ning, YANG Fu-cai, *et al.* Improved SRDCF object tracking via the Best-Buddies similarity[J]. *Optics and Precision Engineering*, 2018, **26**(2): 492-502.
杨德东, 毛宁, 杨福才等, 利用最佳伙伴相似性的改进空间正则化判别相关滤波目标跟踪[J]. 光学精密工程, 2018, **26**(2): 492-502.
- [23] WANG M, LIU Y, HUANG Z. Large margin object tracking with circulant feature maps [C]. IEEE Conference on Computer Vision and Pattern Recognition, Los Alamitos: Washington, DC: IEEE Computer Society Press, 2017: 4800-4808.
- [24] ZHU Z, HUANG G, ZOU W, *et al.* UCT: Learning unified convolutional networks for real-time visual tracking[C]. IEEE International Conference on Computer Vision Workshop, IEEE Computer Society, 2017: 1973-1982.

Foundation item: National Key Research and Development Program of China(No.2017YFC0821102)

引用格式: XIONG Chang-zhen, CHE Man-qiang, WANG Run-ling, *et al.* Adaptive Model Update via Fusing Peak-to-sidelobe Ratio and Mean Frame Difference for Visual Tracking[J]. *Acta Photonica Sinica*, 2018, **47**(9): 0910001

熊昌镇, 车满强, 王润玲, 等. 融合峰旁比和帧差均值自适应模型更新的视觉跟踪[J]. 光子学报, 2018, **47**(9): 0910001