

基于前景感知的时空相关滤波跟踪算法

虞跃洋^{1,2,3,4,5*}, 史泽林^{1,2,3,4,5}, 刘云鹏^{2,3,4,5}

¹ 中国科学技术大学信息科学技术学院, 安徽 合肥 230026;

² 中国科学院沈阳自动化研究所, 辽宁 沈阳 110016;

³ 中国科学院机器人与智能制造创新研究院, 辽宁 沈阳 110016;

⁴ 中国科学院光电信息处理重点实验室, 辽宁 沈阳 110016;

⁵ 辽宁省图像理解与视觉计算重点实验室, 辽宁 沈阳 110016

摘要 针对长时目标跟踪中目标背景混杂、遮挡、目标移出视野导致的跟踪失败问题, 基于空间正则化相关滤波 (SRDCF), 提出一个基于前景感知的时空相关滤波算法。首先, 提出前景感知相关滤波方法, 使得滤波器能够准确地把目标前景区域和背景区域进行区分; 然后, 把前景感知滤波器加入时间正则项中, 使具有时空正则化功能的滤波器始终保持在低维的判别流形上; 同时, 采用交替方向乘子法 (ADMM) 求解, 使得跟踪方法在传统特征的表达上能实现实时性; 最后, 确定目标重检测器的激活阈值, 利用候选区域方法结合相关滤波方法实现重检测, 达到长时跟踪的目的。在标准数据集 OTB-2013 上分别利用传统特征和卷积特征进行实验, 并与 SRDCF 相比, 跟踪平均成功率分别提高了 5.6% 和 7%。本文算法针对目标背景模糊、旋转、遮挡和移出视野等情况, 具有较强的稳健性。

关键词 机器视觉; 目标跟踪; 相关滤波; 时间一致性; 重检测

中图分类号 TP391

文献标识码 A

doi: 10.3788/LOP56.221503

Foreground-Aware Based Spatiotemporal Correlation Filter Tracking Algorithm

Yu Yueyang^{1,2,3,4,5*}, Shi Zelin^{1,2,3,4,5}, Liu Yunpeng^{2,3,4,5}

¹ School of Information Science and Technology, University of Science and Technology of China, Hefei, Anhui 230026, China;

² Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China;

³ Institutes for Robotics and Intelligent Manufacturing, Chinese Academy of Sciences, Shenyang, Liaoning 110016, China;

⁴ Key Laboratory of Opto-Electronic Information Processing, Chinese Academy of Science, Shenyang, Liaoning 110016, China;

⁵ Key Laboratory of Image Understanding and Computer Vision, Shenyang, Liaoning 110016, China

Abstract In this study, we propose a foreground-aware based spatiotemporal correlation filter algorithm based on the spatially regularized discriminative correlation filter (SRDCF) to deal with long-term object tracking failures caused by background clutter, occlusions, and out-of-view objects. Initially, a foreground-aware correlation filtering algorithm is proposed to distinguish the foreground and background of the object accurately. Subsequently, the foreground-aware filter is added to the time regularization term to keep the filter with spatiotemporal regularization function in a low-dimensional discriminative manifold. Simultaneously, the solution based on the alternating direction method of multipliers (ADMM) is conducted to achieve real-time operation of the tracking method in the traditional feature expression. Finally, the activation threshold of object re-detector is determined, and the candidate

收稿日期: 2019-04-01; 修回日期: 2019-04-25; 录用日期: 2019-05-17

基金项目: 中国科学院国防科技创新重点基金 (Y8K4160401)

* E-mail: yuyueyang@sia.cn

region method combined with correlation filtering method is used to achieve re-detection, so as to achieve the purpose of long-term tracking. We conduct experiments using traditional and convolutional features with respect to the OTB2013 standard dataset and observe that the average success rates of tracking are 5.6% and 7% higher, respectively, when compared with that of SRDCF. Therefore, the proposed approach is a robust method for handling background blur, rotations, occlusions, and out-of-view objects.

Key words machine vision; object tracking; correlation filter; temporal consistency; re-detection

OCIS codes 150.0155; 150.1135; 150.1488

1 引言

目标跟踪是计算机视觉中具有挑战性的任务之一。它是一个首先给定视频序列初始帧目标的位置和大小,并在随后的视频序列中估计出目标状态的过程^[1]。近年来,该领域的研究取得了重大进展,但是受到目标外观形态、尺度变化,光照变化和遮挡等问题的影响,目标跟踪任务仍然不能通过任意一个单一跟踪算法处理所有场景。根据目标外观模型的表达策略划分,可将跟踪算法分为生成式模型^[2-3]和判别式模型^[4-8]。生成式模型利用一个模式表达目标的外观形态,然后在搜索图像区域中找到与模型最相似的区域作为目标。判别式模型则是利用已知的样本进行训练,进而学习到一个分类器,从而建立一个将目标从背景区域分离出来的判别方法。而基于相关滤波器的跟踪算法,是判别式模型中的一种,近些年来在各个数据集和目标跟踪的比赛中获得了优秀的效果。

Henriques 等^[9]提出高速核相关滤波算法(KCF),它利用循环密集采样,对目标和目标周围的环境构造出大量训练样本,以提高分类器的分类能力,再利用循环矩阵对角化的性质和快速傅里叶变换,将空域内的计算转换到频域,能降低计算复杂度,实现在线跟踪。Danelljan 等^[10]提出精确尺度估计的稳健跟踪算法(DSSST),其利用多尺度相应策略,能够有效地解决尺度变化的问题。后来, Danelljan 等^[11]提出空间正则化相关滤波跟踪算法(SRDCF),通过引入空间正则化权重来抑制相关响应中背景部分引起的杂波,能降低相关滤波算法中的边界效应,显著提高了相关滤波的跟踪准确性。以 SRDCF 为基础改进的连续空间卷积跟踪器(CCOT)^[12]和高效时空卷积跟踪器(ECO)^[13]也均在公用数据集上表现出卓越的性能。

SRDCF 结合深度神经网络提取的特征,能很好地解决目标外观形态、尺度和光照变化等问题,但是缺乏对跟踪序列中上下文关系的分析,会导致跟踪漂移,尤其是对背景混杂、遮挡、目标移出视野再移

回的情况解决不佳。此外,SRDCF 在传统特征的表达下,计算速度仅为 6 frame/s,远不能达到实时跟踪的目的。为解决这些问题,本文在 SRDCF 的基础上提出一个考虑时间上下文信息的跟踪方法,它利用一个保有历史信息的正则项,使各帧学习到的滤波器始终保持在一个低维流形上,能避免滤波器模板向量突变。在时间正则项的约束中,设计一个具有前景感知功能的滤波方法,其能够利用真实样本训练相关滤波器,避免了循环移位产生的边界效应问题,并能有效区分前景和背景。本文利用交替方向乘子法(ADMM)求解,在几次迭代后即可实现收敛,从而达到快速求解的目的,以实现跟踪的实时性。此外,本文还提出基于目标候选区域的重检测方法,能够有效解决目标发生遮挡、目标移出视野的问题。利用 OTB-2013、OTB-2015、VOT2016 数据集,将本文算法与基准算法 SRDCF 及一些流行的目标跟踪算法进行对比,证明了本文算法的可靠性。

2 空间正则化相关滤波跟踪算法

SRDCF 是一种在频域利用稀疏的正则化矩阵学习得到空间正则化相关滤波器的方法,其有效地减轻了多通道相关滤波器的边界效应,也在各评估基准中证明具有很好的跟踪性能。

在相关滤波的学习过程中,SRDCF 利用空间正则化权重函数 w 调节相关滤波器的 f 。正则化权重函数 w 根据空间位置决定相关滤波器参数 f 的权重大小,即背景区域赋予较高的权重,目标区域赋予较小的权重,来抑制由搜索框增大引起的背景杂波问题。

$$\operatorname{argmin}_f \sum_{k=1}^{k'} \alpha_k \left\| \sum_{d=1}^{d'} \mathbf{x}_k^{(d)} * \mathbf{f}^{(d)} - \mathbf{y}_k \right\|^2 + \sum_{d=1}^{d'} \left\| \mathbf{w} \cdot \mathbf{f}^{(d)} \right\|^2, \quad (1)$$

式中, $\mathbf{x}_k^{(d)}$ 为所提取的多通道特征, $\mathbf{f}^{(d)}$ 为滤波器的值, α_k 决定每个样本 x_k 对相关滤波器参数 f 的影响, $\sum_{d=1}^{d'} \mathbf{x}_k^{(d)} * \mathbf{f}^{(d)}$ 为相关滤波器对样本 x_k 的相关响应, d 为特征维数, d' 为特征总维数, \mathbf{y}_k 为样本 x_k

的软标签, k 为样本帧数, k' 为样本总帧数, $\|\cdot\|$ 为矩阵的 2 范数, $*$ 为相关运算。在这里, Danelljan 等使用高斯-赛德尔方法来迭代更新相关滤波器 f 。这种方法在有效扩大空间搜索区域的同时, 能抑制边界效应。

3 基于前景感知的时空相关滤波跟踪算法

3.1 基于前景感知的时间一致性模型

为表达跟踪过程中时间信息的一致性, 假设构成连续视频帧的滤波器模板的向量位于一个低维流形上, 前后帧变化被限定在一个 l_0 范数中, 即 $\|f - f_{\text{model}}\|_0 < \epsilon$, 其中 f_{model} 为前一帧所求滤波器模板, ϵ 为限定的 l_0 范数值。这种时间一致性的正则项使原 SRDCF 所求得的滤波器模型不会发生跳变, 从而避免跟踪漂移。 l_0 范数非凸, 且其凸包络 l_1 范数被广泛应用于矩阵稀疏度的逼近, 用 l_1 范数松弛方法来获得具有目标区域选择功能的时间一致性模型。将 l_1 范数正则项引入 SRDCF 可得

$$\operatorname{argmin}_f \frac{1}{2} \left\| \sum_{d=1}^{d'} \mathbf{x}_k^{(d)} * \mathbf{f}^{(d)} - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{d=1}^{d'} \left\| \mathbf{w} \cdot \mathbf{f}^{(d)} \right\|^2 + \frac{\mu}{2} \left\| \mathbf{f} - \mathbf{f}_{\text{model}} \right\|_1, \quad (2)$$

式中, \mathbf{y} 为期望的响应, μ 为时间正则项系数。时间一致性正则项增强了每一帧所求当前帧的 f 和历史帧的 f_{model} 的相似性, 同时也能提升 f 的稀疏性。由实验知 μ 的设置远大于空间正则项系数, 并且有 $\|f - f_{\text{model}}\|_2 \leq \|f - f_{\text{model}}\|_1 \leq n \|f - f_{\text{model}}\|_2$, 其中 n 代表 n 维赋范空间, 因此对最优函数放缩可使最优函数简化为

$$\operatorname{argmin}_f \frac{1}{2} \left\| \sum_{d=1}^{d'} \mathbf{x}_k^{(d)} * \mathbf{f}^{(d)} - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{d=1}^{d'} \left\| \mathbf{w} \cdot \mathbf{f}^{(d)} \right\|^2 + \frac{\mu}{2} \left\| \mathbf{f} - \mathbf{f}_{\text{model}} \right\|_2^2. \quad (3)$$

图 1 为本文算法整体流程示意图。和传统的相关滤波方法类似, 滤波器在每一帧进行更新。本文算法在时间正则项中考虑了当前帧学习和历史帧模型的联系, 因此本文算法可以看作是 SRDCF 通过多帧样本学习的合理近似。而且, 在目标发生遮挡或移出视野的情况发生时, SRDCF 会因采集到最近几帧被污染的模型而产生过拟合, 本文方法能通过正则项来约束它与前一帧类似。因为判别式学习的效果和所提供的有监督的样本有密切关系, 所以作为时间一致性约束, 选择训练 f_{model} 的样本至关重要。在这里, 通过掩模矩阵切割将前景与背景分割开, 得到只在目标区域激活的滤波器模板。

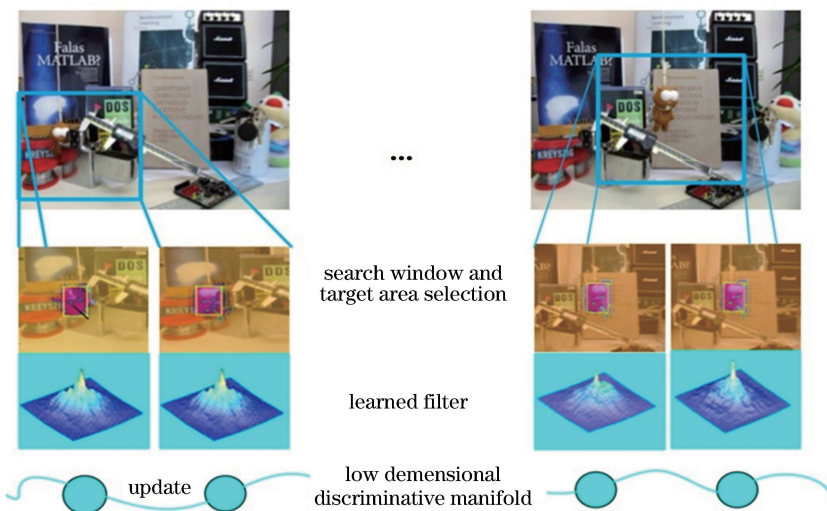


图 1 用序列 Tiger 说明具有目标区域选择功能的时间一致性约束

Fig. 1 Temporal consistency constraints with object area selection function explained by sequence Tiger

3.2 前景感知滤波器模板的学习

本节给出前景感知功能的滤波器模板 f_{model} 的求解。相关滤波跟踪通常使用循环矩阵对角化的性质来对样本特征进行循环移位, 从而产生正负样本学习滤波器。而通过循环移位和密集采样产生的多个样本并不能真实地反应背景和目标区

域, 会造成边界效应。本文用预定义的掩模矩阵, 把目标区域的特征在空域中切割出来作为训练样本, 以区分背景与前景。图 2 为两种方法训练样本的对比。

传统的相关滤波器在频域内求解, 岭回归方程在空域内可以表示为

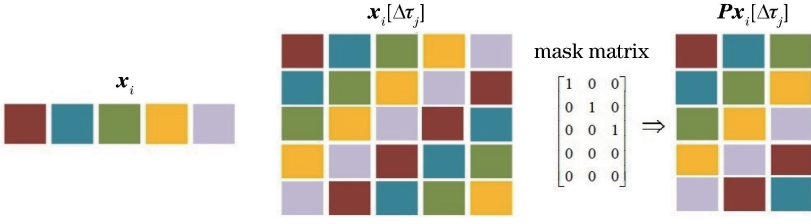


图2 以一维向量为例。假设目标大小 $D=3$ 。左侧是一个 $L=5$ 的一维信号 \mathbf{x}_i , $\mathbf{x}_i[\Delta\tau_j]$ 图像是所有循环移位所得的结果,它通过左乘掩模矩阵 \mathbf{P} ,可以得到 5 个长度为 3 的一维向量,其中前 3 行是与目标大小一致的真实正样本

Fig. 2 Take one-dimensional vector as example, assuming length of target is $D=3$. Left side is one-dimensional signal \mathbf{x}_i with $L=5$. $\mathbf{x}_i[\Delta\tau_j]$ image is result of all cyclic shifts. Five one-dimensional vectors with length of 3 can be obtained by multiplying mask matrix \mathbf{P} on this image, where first 3 rows are real positive samples with same size of object

$$E(\mathbf{f}) = \sum_{d=1}^{d'} \left\| \mathbf{y}_j - \sum_{k=1}^{k'} \mathbf{f}_k^{(d)\top} \mathbf{x}_k[\Delta\tau_j] \right\|_2^2 + \lambda \sum_{k=1}^{k'} \left\| \mathbf{f}_k^{(d)} \right\|_2^2, \quad (4)$$

式中, $E(\mathbf{f})$ 为目标函数, λ 为传统相关滤波器的正则项系数, \mathbf{f}_k 为求取的多通道滤波器, $\mathbf{y}(j)$ 为响应 \mathbf{y} 的第 j 个元素, $[\Delta\tau_j]$ 为循环移位表示符号, $\mathbf{x}_k[\Delta\tau_j]$ 为向量 \mathbf{x}_k 在空域内的第 j 步离散位移, \top 为共轭转置。通过一个二值的掩模矩阵,对当前帧的整幅图像进行分割,把实际样本所在矩形位置作为正样本,把其他和目标区域大小一致的背景切割成块作为负样本进行训练,从而得到只有前景区域的最大响应。并以此作为时间正则项的约束,使其具有前景感知的功能。

定义多个 $L \times D$ 维的二值矩阵 \mathbf{P} , 在整幅输入图像上重新定义(4)式可得

$$E(\mathbf{f}) = \sum_{j=1}^{j'} \left\| \mathbf{y}_j - \sum_{k=1}^{k'} \mathbf{f}_k^\top \mathbf{P} \mathbf{x}_k[\Delta\tau_j] \right\|_2^2 + \lambda \sum_{k=1}^{k'} \left\| \mathbf{f}_k \right\|_2^2. \quad (5)$$

式中, j' 为响应 \mathbf{y} 的元素总数。掩模矩阵 \mathbf{P} 可以把整幅图像分割成目标框大小一致的图像,对整幅图像进行循环移位。这样使得训练样本的数量扩增数

倍,包含了更多的前景背景区分,从而减小过拟合。

由移位所产生真实样本的比例由原来的 $\frac{1}{D}$ 变为现在的 $\frac{L-D+1}{L}$ 。

和传统相关滤波器的求解方法一样,将(5)式转换到频域进行分析。由帕塞瓦尔定理,可得

$$E(\mathbf{f}) = \sum_{j=1}^{j'} \left\| \hat{\mathbf{y}}_j - \text{diag}(\hat{\mathbf{x}}_j)^\top \sqrt{D} \mathbf{F} \mathbf{P}^\top \mathbf{f} \right\|_2^2 + \lambda \left\| \mathbf{f} \right\|_2^2, \quad (6)$$

式中, $\hat{\mathbf{y}}_j$ 为期望响应的傅里叶变换, $\hat{\mathbf{x}}_j$ 为特征的傅里叶变换, $\text{diag}(\cdot)$ 为对角矩阵, \mathbf{F} 为一个标准正交化的正交基向量。通过 \mathbf{F} , 一个 D_1 维的矢量信号可表示为傅里叶变换形式。引入一个等式约束条件后,可以通过增广拉格朗日方法(ALM)求解。由这种方法求得的时间一致性正则项中的滤波器 $\mathbf{f}_{\text{model}}$ 的活跃区域,与目标前景大小一致,能排除边界效应。在后续帧的更新中,ALM 可以作为时间一致性的约束,把前景和背景有效区分。图3为传统相关滤波器和本文训练样本的对比示意图。

3.3 优化方法

SRDCF 采用高斯-赛德尔迭代法求解,高斯-赛

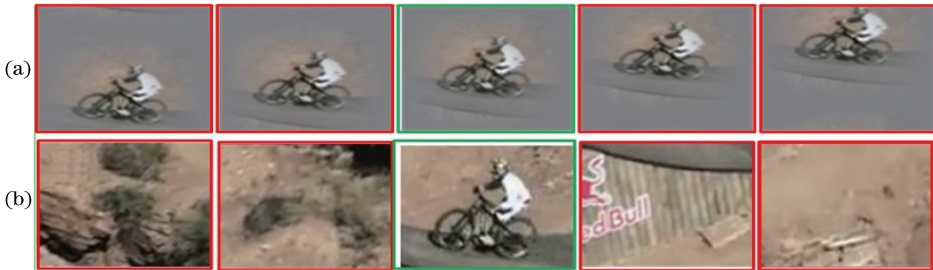


图3 传统相关滤波器和本文方法训练样本对比。(a)传统相关滤波器的循环移位训练样本;
(b)前景感知相关滤波器的训练样本

Fig. 3 Comparison of training samples between traditional correlation filters and proposed method. (a) Cyclic-shift training samples of traditional correlation filter; (b) training samples of foreground-aware correlation filter

德尔迭代法是数值代数中的一种迭代法,用于求解线性方程组,但其收敛速度慢,无法满足跟踪过程的实时性。在这里,采用 ADMM 求解,可以保证在传统特征表达上的跟踪具有实时性。

第 1 帧的滤波器模型通过 3.2 小节求得,为求解(4)式,引入松弛变量 \mathbf{g} ,基于凸优化问题构造的目标函数为

$$\begin{cases} \operatorname{argmin}_{\mathbf{f}, \mathbf{g}} \frac{1}{2} \left\| \sum_{d=1}^{d'} \mathbf{x}_t^{(d)} * \mathbf{f}^{(d)} - \mathbf{y} \right\|^2 + \\ \frac{1}{2} \sum_{d=1}^{d'} \left\| \mathbf{w} \cdot \mathbf{g}^{(d)} \right\|^2 + \frac{\mu}{2} \left\| \mathbf{f} - \mathbf{f}_{\text{model}} \right\|_2^2, & (7) \\ \text{s.t. } \mathbf{g} = \mathbf{f} \end{cases}$$

式中, $\mathbf{g}^{(d)}$ 为多通道滤波器的松弛变量, $\mathbf{x}_t^{(d)}$ 是第 t 帧时刻第 d 个通道的特征。

利用 ALM,将(7)式约束转换为标准函数,则目标函数可以转换为拉格朗日函数的最小值,即

$$\begin{aligned} \mathcal{L}(\mathbf{f}, \mathbf{g}, \zeta) = & \frac{1}{2} \left\| \sum_{d=1}^{d'} \mathbf{x}_t^{(d)} * \mathbf{f}^{(d)} - \mathbf{y} \right\|^2 + \frac{1}{2} \sum_{d=1}^{d'} \cdot \\ & \left\| \mathbf{w} \cdot \mathbf{g}^{(d)} \right\|^2 + \zeta^T \sum_{d=1}^{d'} (\mathbf{f}^{(d)} - \mathbf{g}^{(d)}) + \frac{\gamma}{2} \sum_{d=1}^{d'} \cdot \\ & \left\| \mathbf{f}^{(d)} - \mathbf{g}^{(d)} \right\|_2^2 + \frac{\mu}{2} \left\| \mathbf{f} - \mathbf{f}_{\text{model}} \right\|_2^2, \quad (8) \end{aligned}$$

式中, $\mathcal{L}(\cdot)$ 为拉格朗日函数, ζ 为惩罚因子, γ 为 ALM 的正则项系数。

由于(8)式无闭环解,通过 ADMM 优化求解可得两个优化子问题。

子问题 \mathbf{f} 的求解为

$$\begin{aligned} \operatorname{argmin}_{\hat{\mathbf{f}}} & \left\| \sum_{d=1}^{d'} \hat{\mathbf{x}}_t^{(d)} \cdot \hat{\mathbf{f}}^{(d)} - \hat{\mathbf{y}} \right\|^2 + \\ & \zeta^T \sum_{d=1}^{d'} (\mathbf{f}^{(d)} - \mathbf{g}^{(d)}) + \\ & \frac{\gamma}{2} \sum_{d=1}^{d'} \left\| \mathbf{f}^{(d)} - \mathbf{g}^{(d)} \right\|_2^2 + \frac{\mu}{2} \left\| \mathbf{f} - \mathbf{f}_{\text{model}} \right\|_2^2, \quad (9) \end{aligned}$$

式中, $\hat{\mathbf{x}}_t^{(d)}$ 为多通道特征的傅里叶变换, $\hat{\mathbf{f}}^{(d)}$ 为多通道滤波器的傅里叶变换, $\hat{\mathbf{f}}$ 为 \mathbf{f} 的离散傅里叶变换。已知标签 $\hat{\mathbf{y}}$ 的第 j 个元素仅仅取决于滤波器第 j 个元素的值和样本 $\hat{\mathbf{x}}_t$ 的所有通道。定义 $\mathbf{V}_j(\mathbf{f}) \in \mathbb{R}^{D_2}, \mathbb{R}^{D_2}$ 为由 D_2 维矩阵构成的实数空间。 \mathbf{f} 在 D_2 通道上的第 j 个向量,可以通过向量 $\mathbf{V}_j(\cdot)$ 表示,对(9)式求导并令导数等于 0。可求得子问题的解为

$$\mathbf{V}_j(\hat{\mathbf{f}}) = [\mathbf{V}_j(\hat{\mathbf{x}}_t) \mathbf{V}_j(\hat{\mathbf{x}}_t)^T + (\mu + \gamma) \mathbf{I}]^{-1} [\mathbf{V}_j(\hat{\mathbf{x}}_t) \hat{\mathbf{y}}_j + \gamma \mathbf{V}_j(\hat{\mathbf{g}}_t) - \zeta^T + \mu \mathbf{V}_j(\hat{\mathbf{f}}_{\text{model}})], \quad (10)$$

式中, \mathbf{I} 为单位矩阵, $\hat{\mathbf{g}}_t$ 为松弛变量的傅里叶变换。

子问题 \mathbf{g} 因为没有涉及空域上的卷积运算,因此直接对子问题求导可得空域上的闭环解为

$$\mathbf{g} = (\mathbf{w}^T \mathbf{w} + \gamma \mathbf{I})^{-1} (\zeta^T + \gamma \mathbf{f}). \quad (11)$$

更新拉格朗日系数为

$$\zeta^{(k+1)} = \zeta^{(k)} + \mu [\mathbf{f}^{(k+1)} - \mathbf{g}^{(k+1)}]. \quad (12)$$

更新惩罚系数 γ 为

$$\gamma^{(i+1)} = \min[\gamma_{\max}, \rho \gamma^{(i)}], \quad (13)$$

式中, $\zeta^{(k)}, \zeta^{(k+1)}$ 分别为迭代前后的拉格朗日系数, $\mathbf{f}^{(k+1)}$ 为下一帧迭代求得的滤波器模板, $\mathbf{g}^{(k+1)}$ 为下一次求得的松弛变量, $\gamma^{(i)}, \gamma^{(i+1)}$ 分别为前后迭代的惩罚系数, γ_{\max} 为惩罚系数上限, ρ 为放大倍数。

由 3.1 小节的最优函数可以看出,带有时空正则项的最优函数为凸函数,且 ADMM 的每一个子问题均在空域或频域有闭环解,因此可以收敛到全局最优。

对复杂度分析,由于(9)式的求解在每个像素位置上是相互独立的,因此关于子问题 \mathbf{f} 有 $M \times N$ 个

子问题,每个子问题都含有 D 个变量的线性方程组。每一个线性系统的时间复杂度是 $O(DMN)$,因此解决子问题 $\hat{\mathbf{f}}$ 的复杂度是 $O(DMN)$ 。考虑到离散傅里叶变换和傅里叶逆变换,对滤波器 \mathbf{f} 的求解的复杂度是 $O[DMN \lg(MN)]$ 。对于问题 \mathbf{g} 在空域的闭环解,算法在滤波器求解过程中的复杂度是 $O[DMN \lg(MN) N_1]$,其中 N_1 是 ADMM 的迭代次数。在滤波器求解过程中,算法的复杂度远小于采用高斯-赛德尔方法求解的 $O[(D + K^2) \times DMN N_1]$,大大加快了原优化方法的运算时间。

3.4 重检测

3.4.1 目标重检测

目前任何跟踪算法都不能始终保证跟踪的正确性。基于相关滤波的跟踪器一旦发生跟踪错误,那么所训练到的分类器通常是由前一帧或者前几帧的错误样本取得,那么模板会发生漂移,从而导致跟踪失败。目前常用的办法是用相关响应的峰值旁瓣(PSR)来确定相关峰振荡的剧烈程度,由此来决定跟踪器是否发生跟踪错误。然而,仅仅改变模型跟

新策略无法重捕已丢失的目标。

受目标检测算法的启发,本文采用候选区域和相关滤波结合的方法设计了一个重检测器。文献[14]的候选区域方法为 EdgeBoxes,其利用图像中纹理、边缘、颜色等信息,可以保证在选取较少窗口(几千甚至几百个)的情况下,保持较高的召回率。对于重检测器,本文通过上文方法,获得两个相关滤波器模板。一个与传统相关滤波跟踪方法相同,采用帧帧更新的策略,用于正常跟踪状态;另外一个通过保守的更新率来获得长时跟踪所需的具有记忆效应的滤波器模板,一旦发生目标丢失的现象,能激活重检测器。在重检测时,用 EdgeBoxes 候选区域方法获得目标可能出现的位置,并计算所有候选框目标与具有记忆效应滤波器的相关输出,定义为 $g(\mathbf{b}_t^{(m)})$ 。 $g(\mathbf{b}_t^{(m)})$ 代表第 t 帧中,第 m 个候选框的输出得分, $\mathbf{b}_t^{(m)}$ 为第 t 帧的第 m 个候选框。本文考虑两个连续帧之间位移变化的连续性,计算所有候选区域与上一帧的中心点距离 $D[\mathbf{b}_t^{(m)}, \mathbf{b}_{t-1}]$, \mathbf{b}_{t-1} 为第 $(t-1)$ 帧候选框的位置, $D[\mathbf{b}_t^{(m)}, \mathbf{b}_{t-1}] = \exp\left[-\frac{1}{2\sigma^2} \left\| [\mathbf{x}_t^{(m)}, \mathbf{y}_t^{(m)}] - (\mathbf{x}_{t-1}, \mathbf{y}_{t-1}) \right\|^2\right]$, σ 为方差, $[\mathbf{x}_t^{(m)}, \mathbf{y}_t^{(m)}]$ 为第 t 帧各勾选框的位置坐标, $(\mathbf{x}_{t-1}, \mathbf{y}_{t-1})$ 为上一帧确定的候选框的位置,把所得

到的各个候选框的输出和最小化空间位置项作为最优函数的两部分。这样能兼顾帧与帧间移动的平滑性和最大响应的关系,防止发生重检测目标位置的突变而造成失败。

$$\begin{cases} \operatorname{argmin}_m g[\mathbf{b}_t^{(m)}] + \alpha D[\mathbf{b}_t^{(m)}, \mathbf{b}_{t-1}] \\ \text{s.t. } g[\mathbf{b}_t^{(m)}] > T_r \end{cases}, \quad (14)$$

式中, α 用于调节检测置信度和目标位置置信度的权重, T_r 为接受阈值。

3.4.2 重检测阈值确定

为不与后续实验数据集重合而产生偶然性,用独立的数据集来验证阈值选取的可靠性。MEEM^[15]跟踪算法提供了 10 个带有不同属性的视频序列,如遮挡、光照变化、低分辨率和目标非刚性形变等。VIVID 数据集提供航拍的移动车辆数据,其中包括红外图像和可见光视频序列,其中车辆经常发生互相遮挡、障碍物遮挡和尺度变换等情况,共计上万帧视频序列。为测试重检测器的激活阈值,首先用无重检测功能的跟踪方法,以便获得跟踪失败的序列。以 MEEM 数据集中的 2 个视频序列为例,图 4 为无重检测器算法下,MEEM 数据集中序列 carRace 和序列 ball 交并比(IoU)和置信度得分的关系,图中方框是跟踪结果。

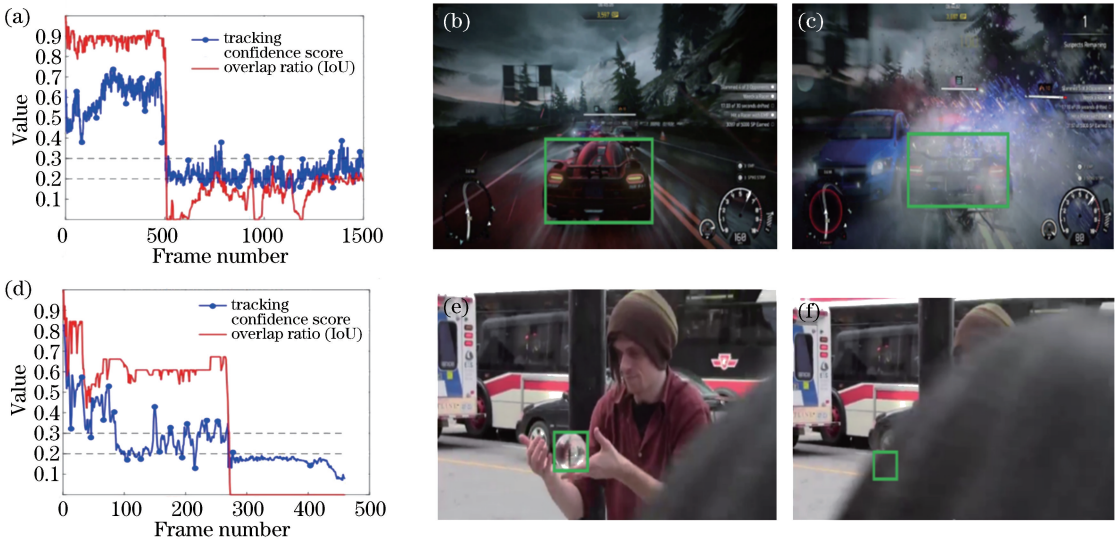


图 4 无重检测器的 carRace 和 ball 序列的 IoU 值和响应得分曲线关系。(a) carRace 的 IoU 值和响应得分曲线关系; (b) carRace 第 502 帧跟踪结果; (c) carRace 第 510 帧跟踪结果; (d) ball 的 IoU 值和响应得分曲线关系; (e) ball 第 209 帧跟踪结果; (f) ball 第 211 帧跟踪结果

Fig. 4 Relationship between IoU value and tracking confidence score for carRace and ball sequences without re-detector. (a) Relationship between IoU value of carRace and tracking confidence score; (b) 502nd-frame tracking result of carRace; (c) 510th-frame tracking result of carRace; (d) relationship between IoU of ball and tracking confidence score; (e) 209th-frame tracking result of ball; (f) 211st-frame tracking result of ball

实验发现,当目标被严重遮挡时,如序列 carRace 的 510 帧和 ball 的 210 帧,相关响应值小于 0.2,会导致跟踪失败(IoU 值小于 0.5),因此将重检测器的阈值 T_0 设置为 0.2。重检测完成以后,需要确定一个接受阈值来判断何时已经重捕目标,并开始正常跟踪。这里用一个比重检测器的激活阈值更大的参数 $K \times T_0$ 来设定这个值,其中 $K \in [1, 2]$, K 是放大系数,并以 0.1 的步长递增,经实验发现,接受阈值取 $1.5T_0$ 时,获得的效果最好。

3.5 算法流程

目标在第 t 帧的定位由前一帧的滤波器 \hat{f}_{t-1} 计算所得。SRDCF 的多尺度处理方法是搜索区域进行一个多分辨率的滤波来估计尺度变化。

搜索区域的大小与滤波器 \hat{f} 在空域的大小一致,结果得到 S 个相关输出,这里 S 为尺度的数目,然后利用插值法找到所有相关输出的最大检测值,有最大相关输出分数的尺度框就可确定目标的尺度和位置。

和传统的相关滤波跟踪器一样,本文采用在线更新策略来提高算法的稳健性,防止过拟合。第 t 帧的滤波器模板以一个固定的学习速率更新, $\hat{f}_{\text{model}} = (1 - \alpha) \hat{f}_{t-1} + \alpha' \hat{f}$,其中 \hat{f}_{model} 为频域的滤波器模板, α' 为在线更新速率。对于第 1 帧的滤波器模板,使用前景感知方法得出,而后对每帧都进行更新,以起到目标区域选择的时间一致性约束作用。

4 实验结果与分析

利用文献[16-17]提供的公开数据集和方法定性、定量分析测试本文算法。在特征表达上用传统特征、基于神经网络提取特征的方法分别进行实验。对本文算法与一些具有代表性的高性能算法进行对比,其中包括 CCOT^[12]、ECO^[13]、DSST^[10]、分层卷积特征跟踪器 CF2^[18]、端到端的相关滤波器跟踪 CFNet^[19]、SRDCF^[11]、KCF^[9]、长时跟踪器 LCT^[4]、多专家模型 MEEM^[15]、孪生全连接网络 SiamFC^[20]、互补学习跟踪方法 Staple^[21]、背景感知跟踪器 BACF^[22]、多域卷积神经网络 MDNet^[23]、多尺度相关滤波跟踪器 SAMF^[24] 及未配备重建测功能的 CF2 跟踪器 CF2camera。最后在文献[15]所提供的数据集中选出序列。在实验数据对比中,本文方法基于传统特征的用 Ours 表示,基于深度特征的用 Oursdeep 表示。实验的测试平台为 Matlab2017a,实验使用的计算机 CPU 为主频

3.40 GHz 的 I7-6700,内存为 8 GB,通过英伟达 GeForce GTX 1080ti GPU 基于 MatConvNet 工具箱提取前向特征。

和 SRDCF 一致,本文在 4 倍于目标长、宽的搜索区域进行查找,然后分别提取方向梯度直方图(HOG)、颜色名称和卷积特征,然后通过余弦窗处理来降低边界效应。对于 ADMM,将(8)式中的超参数 μ 在整个实验中设置为 16。初始的步长参数、最大值和比例因子分别设置为 10、100 和 1.2。ADMM 的迭代次数设置为 2。在第 1 帧滤波器模型的求解中,分别把初始步长、最大值和比例因子分别设置为 1、20 和 5。对于所有实验,固定这些参数不变。

4.1 定量分析

4.1.1 在 OTB-2013 数据集上的评估

利用文献[14]中提供的 2 种评判标准来评价算法:1)精确度,定义为目标中心误差小于指定阈值的帧数占总帧数的百分比,其中阈值取 20 pixel,并把这个值定义为中心距离准确率(DP);2)成功率,定义为 $S = \frac{R_t \cap R_a}{R_t \cup R_a}$,其中 R_t 为跟踪目标的边框区域, R_a 为目标真实的边框区域, \cup 和 \cap 分别为面积的并集和交集运算。当 $S > 0.5$ 时,认为跟踪成功,并把这个值定义为跟踪成功率(OP)。文献[14]中用成功率曲线下的面积(AUC)来确定实际计算成功率的得分。

首先把本文算法和一些其他基于传统特征的先进方法进行对比,表 1 为成功率、精确度、跟踪速度的结果。可以看出,本文算法在平均成功率和阈值精确度方面都比基础算法 SRDCF 高,这是因为引入了时间一致性的约束。ECO 是近年来在各数据集上取得最好成绩的算法之一,其利用高斯混合模型减少了样本的训练维度,但是矩阵的因式分解计算增加了计算复杂度。而本文算法通过 ADMM,经过几次迭代就可实现收敛,避免复杂的数学运算,可以达到实时性(大于 25 frame /s)要求。和 BACF 相比,均把目标区域和背景区域分割作为样本训练分类器的手段,准确率和精确度都有小幅提高。此外,值得注意的是,本文算法的跟踪速度是基准算法 SRDCF 的 4.4 倍。

图 5 为在 OTB-2013 上,基于传统特征的算法的一次通过精度(OPE)曲线和成功率曲线。以成功率曲线和横轴围成的面积表征成功率,本文算法的取值为 0.682,在所有参与比较的先进方法中排名

表 1 基于传统特征的跟踪算法在 OTB-2013 上的成功率、精确度和跟踪速度

Table 1 Success rate, precision, and tracking speed of tracking algorithm based on traditional features on OTB-2013 dataset

| Parameter | Ours | ECO-HC | LCT | SRDCF | Staple-CA | Staple | BACF | DSST | KCF |
|--|------|--------|------|-------|-----------|--------|------|------|-------|
| Mean OP / % | 85.5 | 81.0 | 81.3 | 78.1 | 77.6 | 75.4 | 85.4 | 67.0 | 62.3 |
| Mean DP / % | 89.2 | 87.4 | 84.8 | 83.8 | 83.3 | 79.3 | 78.5 | 74.0 | 74.0 |
| Tracking speed / (frame · s ⁻¹) | 25.3 | 42 | 18.5 | 5.8 | 35.3 | 76.6 | 23.2 | 20.4 | 171.8 |

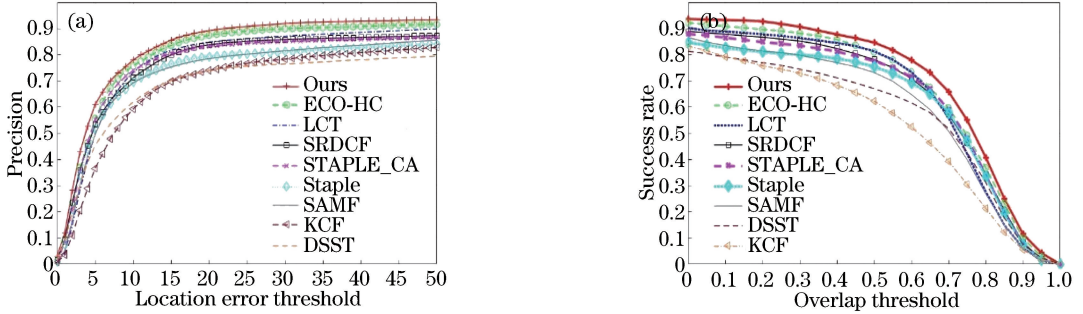


图 5 基于传统特征的跟踪器在 OTB-2013 数据集上的 OPE 曲线和成功率曲线。(a) OPE 曲线;(b)成功率曲线
Fig. 5 Plots of OPE and success rate of trackers with traditional features on OTB-2013 dataset. (a) Plots of OPE;
(b) plots of success rate

第一,其比 SRDCF 和 BACF 分别提高了 5.6% 和 3.2%。本文的方法考虑时间上下文信息,仅把上一帧所得到的滤波器模板作为监督,而同样考虑历史信息 ECO-HC 高效时空卷积跟踪器,把所有的历史信息集合起来,增加了计算复杂度。针对成功率曲线和横轴围成的面积,本文方法比 ECO-HC 提高了 3%。

本文还基于 OTB-2013 数据集对视频序列中的 11 个属性进行了测试。如表 2 所示,本文方法在尺度变换、平面内旋转、背景混杂、照明变化和目标的快速移动 5 项属性中排名第一,其中 SV 为尺度变化,OV 为移出视野,OR 为平面外旋转,OCC 为遮挡,DEF 为形变,MB 为运动模糊,FM 为快速移动,IR 为平面内旋转,BC 为背景混杂,LR 为低分辨率,IV 为照明变化。本文算法的时间一致性正则项包含目标区域的特征选择,能够区分目标和背景区域,从而得到一个较好的跟踪水平。对比基础方法 SRDCF,本文方法在 11 个属性中的性能都有了显著的提高。在遮挡和目标移出视野两个属性中性能分别提高了 11.2% 和 4.2%。这些属性的难点在于目标外观形态上的变化、遮挡和目标移出视野。本文算法增加了时间一致性正则和重检测器,因此比较容易克服这些障碍。

为更好地表征目标物体的模式,本文跟踪器也结合深度特征并与其他基于深度学习的方法进行比较。在这里,对跟踪器组合 VGG-M 中 conv1 和

conv5 的网络输出,结合传统特征作为特征表达,同样采用成功率,精确度和跟踪速度作为衡量标准。表 3 为本文算法和其他基于深度学习方法在 OTB-2013 数据集上的对比。可以看出,在成功率和精确度上,本文算法表现良好,其中成功率比基于卷积特征的 SRDCF 提升了 9.9%,比 VOT2017 比赛季军 ECO 提升了 0.7%。在跟踪速度方面,SiamFC 和 CFNet 都远超实时水平,而本文算法仅仅为 10.6 frame / s,这是由于本文图像在预训练的神经网络中进行前向传播需要经过多次的卷积池化计算来提取特征,这一点与 SiamFC 和 CFNet 相比,处于劣势。针对成功率曲线和横轴围成的面积,如图 6 所示,本文算法以 0.711 排名第一,deepSRDCF 为 0.641,前者比后者提高了 7%。

4.1.2 在 VOT2016 上的评估

表 4 为本文基于传统和深度特征的算法在 VOT2016 上与其他算法的对比。VOT2016 由 60 个具有挑战性的视频序列组成。算法用精确度、稳健性、平均覆盖准确率(EAO)3 个指标进行评估。精确度表示所预测的边界框和实际所标注的物体位置的重合率,稳健性表示在所有序列中,跟踪发生的平均失败次数。EAO 表示对每个跟踪器在一个短时图像序列上的非重叠的期望值。

参与比较的算法有 VOT2015 和 VOT2016 上的冠军算法 MDNet 和 CCOT,由表 4 可以看出,在综合评价指标 EAO 上,本文算法的表现要比

表 2 各跟踪器在 OTB-2013 上的属性评估

Table 2 Performance evaluation of each tracker on OTB-2013 dataset

| Algorithm | SV | OV | OR | OCC | DEF | MB | FM | IR | BC | LR | IV |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| ECO-HC | 0.627 | 0.694 | 0.668 | 0.67 | 0.645 | 0.610 | 0.607 | 0.589 | 0.606 | 0.672 | 0.612 |
| Ours | 0.654 | 0.667 | 0.632 | 0.669 | 0.664 | 0.605 | 0.612 | 0.637 | 0.625 | 0.544 | 0.626 |
| LCT | 0.553 | 0.594 | 0.624 | 0.627 | 0.668 | 0.524 | 0.534 | 0.592 | 0.587 | 0.541 | 0.588 |
| SRDCF | 0.587 | 0.555 | 0.599 | 0.627 | 0.635 | 0.601 | 0.569 | 0.566 | 0.587 | 0.541 | 0.576 |
| SAMF | 0.507 | 0.555 | 0.559 | 0.612 | 0.625 | 0.461 | 0.483 | 0.525 | 0.520 | 0.526 | 0.513 |
| Staple-CA | 0.574 | 0.562 | 0.594 | 0.600 | 0.632 | 0.569 | 0.566 | 0.601 | 0.587 | 0.497 | 0.596 |
| Staple | 0.551 | 0.547 | 0.575 | 0.593 | 0.618 | 0.541 | 0.508 | 0.580 | 0.576 | 0.496 | 0.568 |
| KCF | 0.427 | 0.550 | 0.495 | 0.514 | 0.534 | 0.497 | 0.459 | 0.497 | 0.535 | 0.537 | 0.493 |
| DSST | 0.546 | 0.462 | 0.536 | 0.532 | 0.506 | 0.455 | 0.428 | 0.563 | 0.517 | 0.345 | 0.561 |

表 3 基于卷积特征的跟踪算法在 OTB-2013 上的准确率、精确度和跟踪速度

Table 3 Success rate, precision, and tracking speed of tracking algorithm based on convolutional features on OTB-2013 dataset

| Parameter | Ours | ECO | MDNet | CCOT | DeepSRDCF | SiamFC | CFNet | CF2 |
|---|------|------|-------|------|-----------|--------|-------|------|
| Mean OP / % | 89.4 | 88.7 | 91.1 | 83.2 | 79.5 | 79.1 | 76.9 | 74.0 |
| Mean DP / % | 90.0 | 93.0 | 94.8 | 89.9 | 84.9 | 81.5 | 80.7 | 89.1 |
| Tracking speed / (frame · s ⁻¹) | 10.6 | 9.8 | 0.8 | 0.8 | 0.2 | 83.7 | 78.4 | 10.2 |

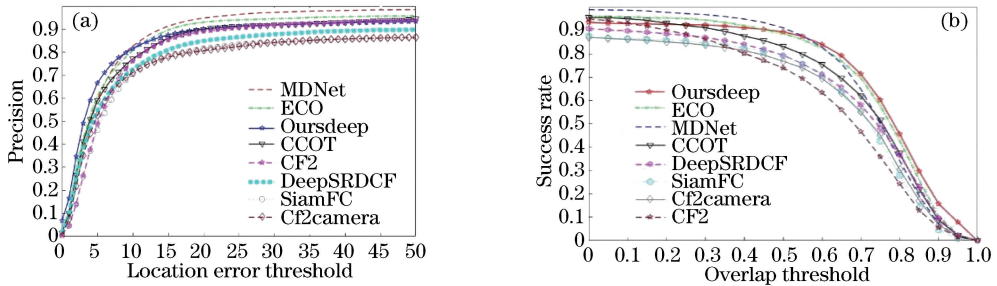


图 6 基于卷积特征的跟踪器在 OTB-2013 数据集上的 OPE 曲线和成功率曲线。(a) OPE 曲线;(b)成功率曲线

Fig. 6 Plots of OPE and success rate of trackers with convolutional features on OTB-2013 dataset. (a) Plots of OPE; (b) plots of success rate

SRDCF 和 BACF 更好。传统特征版本的 Ours 跟踪器的表现,要比具有 CNN 特征的 SRDCF 还要提升 4.4%。

4.2 定性分析

图 7 为本文两种特征表达版本的跟踪方法,在 OTB-2015 数据集中节选 8 个比较有难度的视频序列,并和其他一些高水准算法进行比较。CF2 利用分层卷积特征可以有效地在旋转和目标形变的序列(motorrolling and skiing)中获得成功,但是无法有效抵抗背景混杂和光照变化(singer2 and soccer)。基于相关滤波的跟踪器,如 CF2, KCF, DSST, SRDCF 等,在发生目标遮挡时,滤波器模板迅速污染并发生漂移(girl2)。本文算法配以稳健性更强的卷积特征作为模式表达,能有效抵抗这些困难,因为时间一致性正则项具有前景感知的约束功能,能够

表 4 VOT2016 数据集上各算法的 EAO, 精确度和稳健性评估

Table 4 Evaluations of EAO, precision, and robustness of algorithms on VOT2016 dataset

| Algorithm | EAO | Accuracy | Robustness |
|-----------|-------|----------|------------|
| DSST | 0.181 | 0.500 | 2.720 |
| ECO | 0.375 | 0.530 | 0.730 |
| Staple | 0.295 | 0.540 | 1.350 |
| MDNet | 0.257 | 0.530 | 1.200 |
| BACF | 0.223 | 0.560 | 1.880 |
| SRDCF | 0.247 | 0.520 | 1.500 |
| ECO-HC | 0.322 | 0.510 | 1.080 |
| DeepSRDCF | 0.276 | 0.510 | 1.170 |
| CCOT | 0.331 | 0.530 | 0.238 |
| SiamFC | 0.277 | 0.549 | 0.382 |
| Ours | 0.320 | 0.535 | 0.926 |
| Oursdeep | 0.285 | 0.555 | 1.330 |

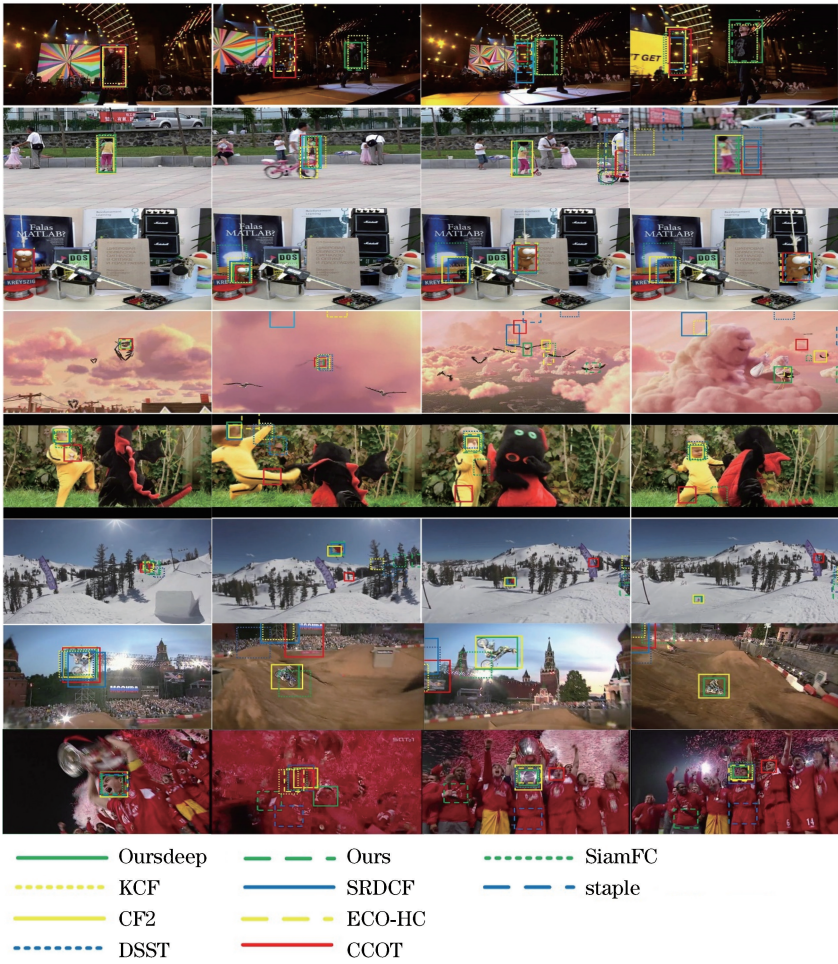


图7 本文算法和 SiamFC,CCOT,DSST,KCF,ECO,CF2 在 OTB-2015 上的 8 个序列的效果对比。从上到下依次为 singer2,girl2,tiger,bird1,dragonbaby,motorrolling,skiing,soccer

Fig. 7 Comparison of tracking results of SiamFC, CCOT, DSST, KCF, ECO, CF2, and proposed algorithm on 8 challenging sequences from OTB-2015 dataset. From top to bottom: singer2, girl2, tiger, bird1, dragonbaby, motorrolling, skiing, and soccer

有效抑制背景与目标的混杂,且正则项可以阻止帧与帧之间滤波器模板因突变而产生的漂移。此外,重检测块能够抵抗遮挡影响。这种时间一致性的约束和重检测器甚至可以在目标发生完全遮挡,移出视野并移回的序列中起到效果(bird1)。从 4.1 小节的属性性能比较中可以看出,本文算法对平面内旋转和平面外旋转场景的处理效果很好(dragonbaby),因为其延续了 SRDCF 空间正则化方法的优点,总是能在较大范围的搜索框中很好地区分目标区域和周围环境区域。

5 结 论

空间正则化相关滤波算法 SRDCF 在目标超出视野、严重遮挡和复杂背景的情况下容易跟踪失败。基于 SRDCF,引入时间一致性约束,使相关滤波跟

踪同时具有时间空间正则化功能。令跟踪模型能够考虑历史信息,在时间正则项中,引入具有前景感知功能的滤波器作为约束,并消除边界效应的影响。在跟踪整体框架上使用 ADMM,能够改变 SRDCF 运算速度慢的缺陷,从而实现在线跟踪。此外,通过设计基于候选区域方法的重检测器,探究重检测器激活的条件,使跟踪器能有效对抗遮挡和移出视野。利用 OTB-2013,OTB-2015 和 VOT2016 数据集,对本文算法和时下最先进的算法进行对比。结果显示,在 OTB-2013 上,本文算法的传统和卷积特征两个版本的跟踪成功率比 SRDCF 均有提高,且与参与比较的算法相比,在跟踪成功率排名上均为第一。通过基于数据集的定性和定量分析,本文算法能够应对多种复杂场景,与现有先进算法相比较,有很好的稳定性。

参 考 文 献

- [1] Gao M F, Zhang X X. Scale adaptive kernel correlation filtering for target tracking[J]. *Laser & Optoelectronics Progress*, 2018, 55(4): 041501.
高美凤, 张晓玄. 尺度自适应核相关滤波目标跟踪[J]. *激光与光电子学进展*, 2018, 55(4): 041501.
- [2] Ross D A, Lim J, Lin R S, *et al.* Incremental learning for robust visual tracking[J]. *International Journal of Computer Vision*, 2008, 77(1): 125-141.
- [3] Kwon J, Lee K M. Visual tracking decomposition [C] // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA. New York: IEEE, 2010: 1269-1276.
- [4] 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.
- [5] Cai Y Z, Yang D D, Mao N, *et al.* Visual tracking algorithm based on adaptive convolutional features [J]. *Acta Optica Sinica*, 2017, 37(3): 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(3): 0315002.
- [6] Hare S, Saffari A, Torr P H S. Struck: structured output tracking with kernels[C] // 2011 International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain. New York: IEEE, 2011: 263-270.
- [7] Gao J, Ling H B, Hu W M, *et al.* Transfer learning based visual tracking with Gaussian processes regression[M] // Fleet D, Pajdla T, Schiele B, *et al.* Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8691: 188-203.
- [8] Liao X F, Hou Z Q, Yu W S, *et al.* A scale adapted tracking algorithm based on kernelized correlation [J]. *Acta Optica Sinica*, 2018, 38(7): 0715002.
廖秀峰, 侯志强, 余旺盛, 等. 基于核相关的尺度自适应视觉跟踪[J]. *光学学报*, 2018, 38(7): 0715002.
- [9] 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.
- [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] 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.
- [12] Danelljan M, Robinson A, Shahbaz Khan F, *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.
- [13] 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.
- [14] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[C] // 2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE, 2013: 2411-2418.
- [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, Pflugfelder R, Leonardis A, *et al.* The visual object tracking vot2016 challenge results[C] // European Conference on Computer Vision Workshops (ECCVW), October 8-10, 2016, Amsterdam, The Netherlands. New York: IEEE, 2016: 777-823.
- [17] 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.
- [18] 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.
- [19] Valmadre J, Bertinetto L, Henriques J, *et al.* End-to-end representation learning for correlation filter

- based tracking [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 5000-5008.
- [20] Zhang J M, Ma S G, Sclaroff S. MEEM: robust tracking via multiple experts using entropy minimization [M] // Fleet D, Pajdla T, Schiele B, *et al.* Computer vision-ECCV 2014. Lecture notes in computer science. Cham: Springer, 2014, 8694: 188-203.
- [21] Bertinetto L, Valmadre J, Henriques J F, *et al.* Fully-convolutional Siamese networks for object tracking [C] // Hua G, Jégou H. Computer vision-ECCV 2016 Workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 850-865.
- [22] 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.
- [23] 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.
- [24] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 4293-4302.