

doi: 10.11830/ISSN.1000-5013.201703017



改进的核相关滤波目标跟踪算法

谢维波, 夏远祥, 刘文

(华侨大学 计算机科学与技术学院, 福建 厦门 361021)

摘要: 提出一种基于核相关滤波的尺度和旋转自适应跟踪算法. 首先, 利用核相关滤波确定目标的中心位置; 然后, 使用特征点匹配的方式估计目标的尺度变化和旋转角度. 在特征点匹配过程中, 使用前、后两次光流匹配消除不稳定特征点; 计算特征点对的权重分布, 从而估计出目标的最佳尺度和角度; 判断当前目标是否受到遮挡, 进而使用更合理的方式更新特征点集和目标模型, 进一步提高算法的鲁棒性. 实验结果表明: 文中算法不仅能在一定程度上处理目标外观变化问题, 而且跟踪的实时性较好.

关键词: 视频跟踪; 核相关滤波; 尺度计算; 角度计算; 遮挡检测

中图分类号: TP 311 **文献标志码:** A **文章编号:** 1000-5013(2017)03-0379-06

Improved Object Tracking Algorithm Based on Kernelized Correlation Filter

XIE Weibo, XIA Yuanxiang, LIU Wen

(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract: A scale and rotation adaptive tracking algorithm based on kernelized correlation filter is proposed. Firstly, the algorithm determines the center position of the object via kernelized correlation filter. Then the algorithm estimates the scale changes and rotation angle of an object using keypoints matching. In the process of keypoints matching, the method eliminates unstable keypoints using forward and backward matching. The next, the algorithm estimates the optimal scale and angle by considering the weight of keypoints. At last, the method detects whether the target is occluded, and then update keypoints set and object model more reasonable, and hence improving the robustness of the algorithm. The results of experiments show that the proposed algorithm not only can handle changes of object's appearance to a certain degree, but also have high tracking efficiency.

Keywords: visual tracking; kernelized correlation filter; scale calculation; angle calculation; occlusion detection

目标跟踪是计算机视觉领域中最具挑战的问题之一, 在视频监控、机器人等领域^[1-9]有着广泛的应用. 近年来, 基于相关滤波的跟踪算法^[1-4]凭借其较快的速度在跟踪领域取得了不错的效果. Henriques 等^[4]提出基于核相关滤波(kernelized correlation filter, KCF)的跟踪算法, 在跟踪目标时有不错的效果. 然而, 当目标尺度发生变化或旋转时, 目标模型不能充分描述目标的表观特征, 而容易跟踪失败. 本文在 KCF 跟踪算法的基础上, 通过引入特征点匹配的方法对目标尺度和角度进行估计, 实现目标的尺度、角度自适应跟踪.

收稿日期: 2016-05-14

通信作者: 谢维波(1964-), 男, 教授, 博士, 主要从事信号处理、视频图像分析的研究. E-mail: xwblxf@hqu.edu.cn.

基金项目: 国家自然科学基金资助项目(61271383); 华侨大学研究生科研创新能力培育计划资助项目(1400214007)

1 核相关滤波模型

传统的相关滤波跟踪算法被视为线性岭回归模型,即

$$\min \sum_i (f(x_i) - y_i)^2 + \lambda \| \mathbf{w} \|^2. \tag{1}$$

式(1)中: x_i 为输入; y_i 为相应的标签; \mathbf{w} 为相应的系数. 这种模型在处理非线性问题上会受到限制,通过使用核方法^[10],可以解决非线性问题. Henriques 等^[4]通过引入岭回归和循环矩阵将相关滤波器核化,则训练模型阶段的表达式为

$$\hat{\boldsymbol{\alpha}} = \hat{\mathbf{y}} / (\hat{\mathbf{k}} + \lambda). \tag{2}$$

式(2)中: $\hat{\boldsymbol{\alpha}}$ 为 \mathbf{w} 的对偶表达; $\hat{\mathbf{k}}$ 为核函数; $\hat{\mathbf{y}}$ 为样本标签. 则检测阶段的表达式为

$$\tilde{\mathbf{y}} \langle \mathbf{w}, \varphi(\mathbf{z}) \rangle = \sum_i \alpha_i \mathbf{k}(\mathbf{x}_i, \mathbf{z}) = \mathbf{C}(\mathbf{K}^{\mathbf{x}, \mathbf{z}}) \boldsymbol{\alpha} = \mathbf{F}^{-1}(\hat{\mathbf{k}}^{\mathbf{x}, \mathbf{z}} \odot \hat{\boldsymbol{\alpha}}). \tag{3}$$

式(3)中: $\langle \cdot \rangle$ 为求内积; \mathbf{x} 为上一帧学习到的目标外观模型;置信输出 $\tilde{\mathbf{y}}$ 的元素值表示对应候选区域检测到目标的概率值,值最大的位置就是目标在新一帧图像中的位置.

2 目标尺度估计和旋转角度估计

根据特征点之间相对距离和相对角度的变化估计目标的尺度和角度,同时,通过计算特征点对的权重分布提高估计的准确度,并提出了一种遮挡检测机制,以提高算法的鲁棒性.

2.1 前、后向光流匹配

根据特征点估计目标的尺度和角度,因此,如何匹配到稳定的特征点就变得尤为重要. 文中算法采用耗时较少 FAST 特征子^[11]检测图像特征点,并使用 256 维的 BRIEF^[12]描述子对其进行描述.

设 $t-1$ 帧时,目标特征点集为 $M^{t-1} = \{x_i^{t-1}\}_{i=1}^n$. 为了得到特征点 x_i^{t-1} 在第 t 帧时的匹配点 x_i^t ,采用耗时较少的 KLT 光流匹配技术^[13]对 $t-1$ 时刻的特征点进行匹配. 为了提高匹配精度,通过前、后两次匹配消除不稳定的特征点.

1) $t-1$ 帧时的特征点 x_i^{t-1} 进行 KLT 光流匹配技术,得到 t 帧时的运动目标特征点 x_i^t ,即为前向匹配;2) 据得到的 t 帧时的目标特征点 x_i^t ,经过金字塔的 KLT 光流匹配技术也能反向推导出前一帧中的目标特征点 x_i^{t-1} ,即为后向匹配;3) 算前、后向匹配特征点描述符的误差距离 $d(x_i^{t-1}, x_i^{t-1})$,当距离大于阈值 D_0 (设为经验值 30) 时,就认为这个特征点是不稳定的特征点,将其前向匹配的特征点 x_i^t 剔除;反之,则保留. 最后,可得到 t 帧时稳定的目标特征点集 $M_t = \{x_i^t\}_{i=1}^m$.

2.2 遮挡检测

在跟踪视频的初始帧中,初始目标框人为选定,则首次匹配的特征点集 M_s 是跟踪目标的稳定描述,称为稳定特征点集. 设 t 帧时匹配的特征点集为 M' ,数量记为 $N_t = |M'|$. 当目标被遮挡时, N_t 将会变得很小. 因此,文中算法根据稳定特征点的数量 $N_s = |M_s|$ 和 t 帧时匹配特征点集的数量 N_t 的比值来确定跟踪目标是否被遮挡,即

$$u_0 = \begin{cases} 1, & N_t/N_s < \theta_u; \\ 0, & \text{其他.} \end{cases} \tag{4}$$

式(4)中:阈值 θ_u 设定为经验值 0.5. 定义 t 帧时活动特征点集 M'_a 用于尺度和角度估计,当目标未被遮挡时, $M'_a = M_t$.

2.3 特征点对的权重分布

在前、后向光流匹配中,特征点描述符的误差距离 $d(x_i^{t-1}, x_i^{t-1})$ 可以衡量特征点 x_i^t 的稳定性(即前、后帧特征点的相似性). 则特征点对 x_i^t 和 x_j^t , t 帧时的稳定性可表示为

$$d_t(i, j) = d(x_i^{t-1}, x_j^{t-1}) + d(x_j^{t-1}, x_i^{t-1}). \tag{5}$$

$d_t(i, j)$ 越小,表示特征点对 x_i^t 和 x_j^t 越稳定,相应地,在估计尺度和角度时,应占更大的权重. 定义 t 帧时特征点对 x_i^t, x_j^t 的权重分布为

$$w_t(i, j) = \frac{k(-\|\frac{d_t(i, j)}{h}\|^2)}{\sum_{1 \leq x < y \leq m} k(-\|\frac{d_t(x, y)}{h}\|^2)}. \quad (6)$$

式(6)中: m 为 t 帧时活动特征点集中特征点的个数; h 为带宽参数; k 为高斯核函数.

2.4 目标最佳尺度估计

在跟踪视频的初始帧中, 给定跟踪目标框 $b_1 = [x_1, y_1, w_1, h_1, \theta_1]$. 其中, x_1, y_1 为目标框的中心位置; w_1 和 h_1 为目标框的宽度和高度; θ_1 为目标框的偏转角度. 在第 t 帧活动点集 $M_a^t = \{x_1^t, x_2^t, \dots, x_m^t\}$, 其中, m 为特征点个数. x_i^t 表示特征点 x_i^{t-1} 在 t 帧中的位置, 则特征点对 x_i^t 和 x_j^t 间距离变化的比例为

$$r_t(i, j) = \|x_i^t - x_j^t\| / \|x_i^{t-1} - x_j^{t-1}\|. \quad (7)$$

式(7)中: $\|\cdot\|$ 为欧氏距离.

考虑特征点对的权重, 则尺度伸缩系数为

$$\bar{r}_t(i, j) = \sum_{1 \leq i < j \leq m} w_t(i, j) r_t(x, y). \quad (8)$$

则 t 时刻的目标框的宽度和高度分别为

$$w_t = \bar{r}_t w_{t-1} = w_1 \prod_{i=2}^t r_i, \quad (9)$$

$$h_t = \bar{r}_t h_{t-1} = h_1 \prod_{i=2}^t r_i. \quad (10)$$

当目标未被遮挡时, 更新目标尺度; 否则, 无需更新. 则目标框的宽度和高度分别为

$$w_t = u_0 w_{t-1} + (1 - u_0) \bar{r}_t = w_{t-1}, \quad (11)$$

$$h_t = u_0 h_{t-1} + (1 - u_0) \bar{r}_t = h_{t-1}. \quad (12)$$

2.5 目标旋转角度估计

当跟踪目标在平面内旋转时, 正确匹配的特征点之间的相对方向也会发生旋转, 并且和目标旋转角度一致. 因此, 文中算法根据前、后两帧特征点之间相对角度的变化估计目标框的旋转角度.

特征点对的角度变化示例, 如图1所示. 设 $t-1$ 帧时, 任意两个特征点之间的角度为 $\arctan(x_i^{t-1}, x_j^{t-1})$, t 帧时相应的角度为 $\arctan(x_i^t, x_j^t)$, 则可得 t 帧时前、后两帧特征点 x_i^t 和 x_j^t 之间变化的角度为

$$\alpha_t(i, j) = \arctan(x_i^t, x_j^t) - \arctan(x_i^{t-1}, x_j^{t-1}). \quad (13)$$

考虑特征点对的权重, 则 t 帧时角度变化为

$$\alpha_t = \sum_{1 \leq i < j \leq m} w_t(i, j) \alpha_t(i, j). \quad (14)$$

式(14)中: m 为活动特征点集中特征点的个数. 在文中算法中, 当

目标被遮挡或角度变化 α_t 小于阈值 α_d 时, 跟踪框的角度变化无需更新. 则 t 帧时目标框的角度为

$$\theta_t = \begin{cases} \theta_{t-1}, & \alpha_t < \alpha_d, \\ u_0 \theta_{t-1} + (1 - u_0)(\theta_{t-1} + \alpha_t), & \text{其他}. \end{cases} \quad (15)$$

实验中, 首帧目标框的角度 θ_1 为 90° .

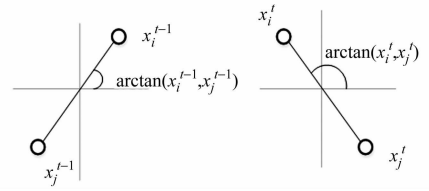


图1 特征点对的角度变化示例

Fig. 1 Examples of pairwise angular change between keypoints

3 文中跟踪算法

3.1 基本原理

3.1.1 确定目标的位置 文中算法将跟踪目标分成确定目标位置和确定目标尺度、角度两部分. 目标状态用矩形框 $b = [x, y, w, h, \theta]$ 表示, 设定目标搜索范围 of 原目标框大小的 2 倍.

根据上述的训练方法, 由训练样本可得 $\hat{\alpha} = \hat{y} / \hat{k}^{x, x'} + \lambda$, 文中算法的核函数采用西不变的高斯核函数

$$\hat{k}^{x, x'} = \exp(-\frac{1}{\sigma^2}(\|x\|^2 + \|x'\|^2 - 2F^{-1}(\hat{x} \odot \hat{x}'))). \quad (16)$$

对于下一帧新的样本 z , 利用训练好的 α , 计算傅里叶域的响应值, 即

$$\hat{\mathbf{y}}(\mathbf{z}) = (\hat{\mathbf{k}}^{\mathbf{x},\mathbf{z}}) \odot \hat{\boldsymbol{\alpha}}.$$

(17)

式(17)中: \mathbf{x} 为学习到的外观模型. 将 $\hat{\mathbf{y}}(\mathbf{z})$ 作逆变换, 转到空间域并求最大值, 即 $\arg \max F^{-1} \hat{\mathbf{y}}(\mathbf{z}^i)$. 其中, 响应值最大的位置即为跟踪目标的中心位置.

3.1.2 确定目标的尺度变化和旋转角度变化 确定目标的中心位置后, 检测并匹配新一帧图像的特征点. 根据式(4)判断目标是否被遮挡, 根据式(9), (10), (14)确定目标框的尺度和旋转角度.

3.1.3 确定更新方案 为了更好地适应跟踪过程中目标外观的变化, 准确估计当前目标的尺度和旋转角度, 需要及时对活动点集、相关滤波模板和目标外观模型进行更新.

a) 动点集更新. 首先, 检测当前目标是否被遮挡, 如果没有发生遮挡, 则用当前匹配的特征点集完全替换上一帧匹配的活动点集; 如果目标被遮挡, 则当前活动点集无需更新, 即

$$M_a^t = u_0 M_a^{t-1} + (1 - u_0) M_t.$$

(18)

式(18)中: M_t 为和 M_a^{t-1} 分别表示 t 帧时匹配的特征点集和 $t-1$ 帧时的活动点集.

b) 波模型和外观模型更新. 传统的 KCF 算法采用固定模式更新滤波模板和目标外观模型, 但当目标被遮挡时, 这种更新方式会带来误差. 为了提高跟踪精度, 当目标被遮挡时, 停止对滤波模板和外观模型的更新, 即

$$\hat{\boldsymbol{\alpha}}_{t+1} = u_0 \hat{\boldsymbol{\alpha}}_{t-1} + ((1 - u_0)(1 - \beta) \hat{\boldsymbol{\alpha}}_{t+1} + \hat{\boldsymbol{\beta}}_t),$$

(19)

$$\hat{\mathbf{x}}_{t+1} = u_0 \hat{\mathbf{x}}_{t-1} + ((1 - u_0)(1 - \beta) \hat{\mathbf{x}}_{t+1} + \hat{\boldsymbol{\beta}}_t).$$

(20)

式(20)中: β 为模型更新因子.

文中算法的流程图, 如图 2 所示.

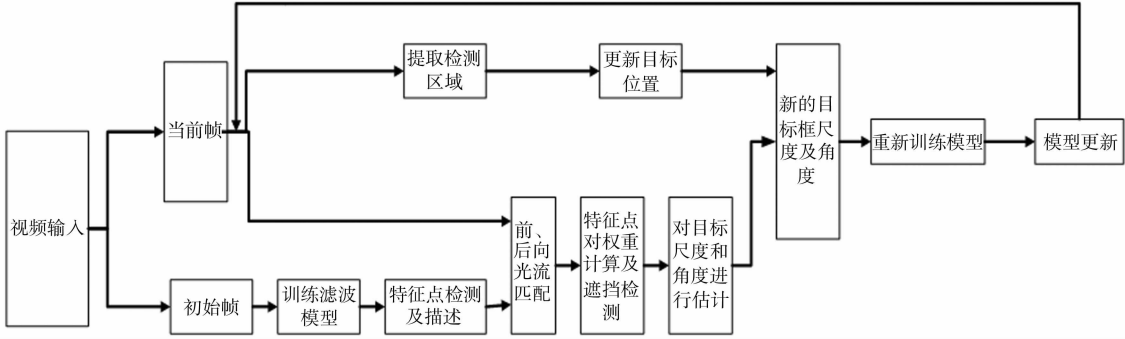


图 2 文中算法的流程图
Fig. 2 Flow chart of algorithm

4 实验结果分析

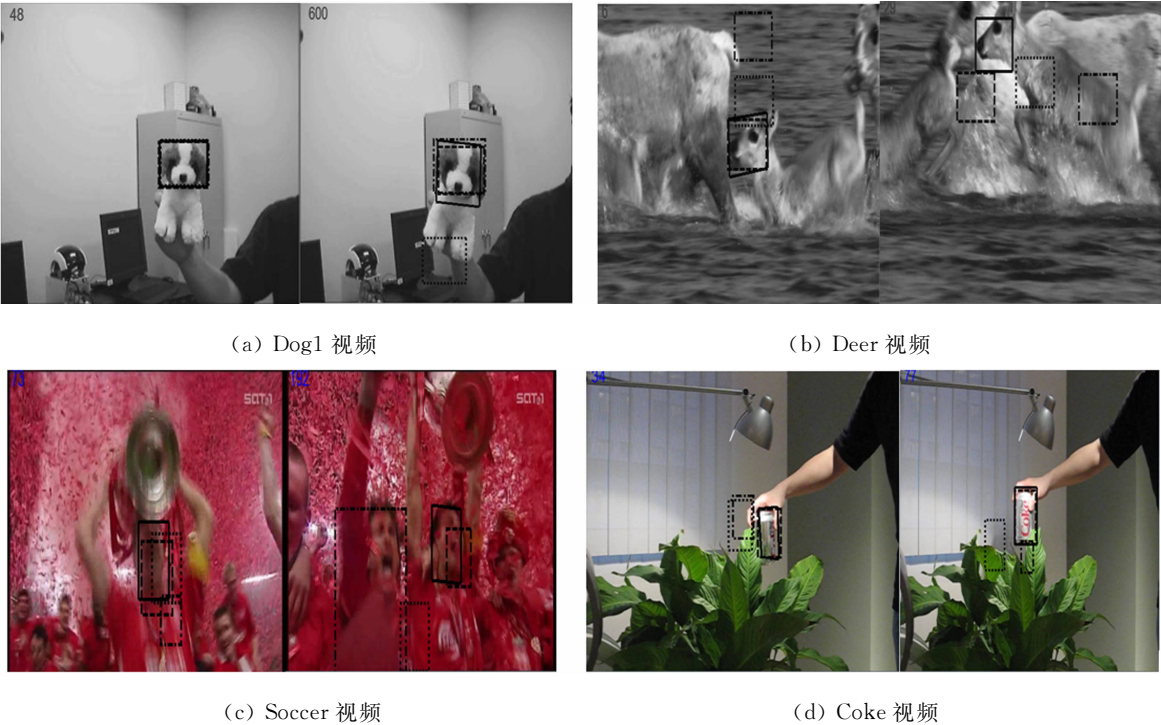
为了对文中算法性能进行更加准确的评估, 使用文献[7]中的 10 组目标尺度变化和旋转变化的标准视频序列对算法进行测试. 同时, 与当前主流的跟踪算法 DFT^[5], STC^[3] 和基准方法 KCF 进行对比. 实验中, 对比算法来自相应作者的公开代码, 并且保留默认的参数设置. 文中算法的正则化参数 λ 和带宽参数 h 设置为 0.01, 角度变化阈值 α_d 设为 2, 模型更新因子 β 设为 0.075. 所有实验都在 Windows 7 系统, Core i5-3470 3.20 GHz CPU 和 8 GB RAM 的台式机上实现, 文中算法在 Matlab 2012 和 Visual Studio 2010 平台上实现.

4.1 定性分析

文中算法和对比算法在部分视频序列上的部分跟踪结果, 如图 3 所示. 这些视频序列不仅有尺度和旋转问题, 还有遮挡(Soccer 和 Coke)、光照(Dog1, Soccer 和 Coke)等问题. 不同算法的跟踪结果用不同线型的矩形框表示, 其中, 实线框表示文中算法, 线段框表示 KCF 算法, 点框表示 DFT 算法, 点线框表示 STC 算法.

Dog1 视频的部分跟踪结果, 如图 3(a)所示. 由图 3(a)可知: 在第 600 帧时, 由于跟踪目标由正面转为侧面, 目标姿势及相对大小都发生了一定的改变, 此时, DFT 算法不能很好地跟踪目标, 其他算法能跟踪目标, 而只有文中算法的跟踪框的角度能随之改变. Deer 视频的部分跟踪结果, 如图 3(b)所示. 由

图 3(b)可知:在第 6 帧时,由于跟踪目标骤降,DFT,STC 跟踪失败;第 29 帧时,KCF 跟踪失败,文中算法一直有着较好的跟踪效果. Soccer 视频的部分跟踪结果,如图 3(c)所示. 由图 3(c)可知:在第 73 帧时,跟踪目标上下剧烈运动,STC 算法产生了较大的跟踪偏差;在第 192 帧时,产生了较大的背景干扰,同时,目标发生了有角度的运动,相对大小也发生了改变,KCF 算法逐渐丢失跟踪目标,文中算法仍然有着较好的跟踪效果. Coke 视频的部分跟踪结果,如图 3(d)所示. 由图 3(d)可知:在第 34 帧时,目标由上到下时,产生了部分遮挡,同时,光线由明转暗且产生了有角度的旋转,DFT 产生了较大的跟踪偏差;第 77 帧时,目标由近至远,尺度发生了变化时,KCF 和文中算法仍能很好地跟踪目标. 整体看来,文中算法有着较好的跟踪精度.



(a) Dog1 视频 (b) Deer 视频

(c) Soccer 视频 (d) Coke 视频

图 3 跟踪算法的性能比较

Fig. 3 Performance comparison of tracking algorithms

4.2 定量分析

为了验证文中算法的有效性,用帧率、中心位置误差(centre location error,CLE)和距离精度(distance precision,DP)3 个指标对比其他跟踪算法. 其中,帧率用来衡量跟踪速度;平均中心位置误差及距离精度用来衡量跟踪结果的准确度.

4.2.1 跟踪速度 帧率表示每秒运行的帧数,值越大跟踪速度越快. DFT,STC,KCF 和文中算法的帧率分别为:13.1,397.4,39.4,79.3 帧 · s⁻¹. 由此可知:虽然需要对尺度和旋转角度进行估计,但文中算法在测试视频序列中仍有较好的跟踪实时性(每秒运行帧数达到 15~20,即可视为实时跟踪).

4.2.2 跟踪精度 中心位置误差 CLE 表示跟踪目标中心位置与标注目标中心位置之间的平均欧式距离,其值越小,表示跟踪精度越高. 各算法在视频序列上的目标中心位置误差,如表 1 所示.

由表 1 可知:文中算法比其他算法有较低的 CLE,在平均 CLE 上取得了最优的效果. 与 KCF 算法相比,文中算法的 CLE 平均降低了 8.39 px.

距离精度 DP 定义为 CLE 小于某一阈值的帧数占视频总帧数的百分比. 设定阈值为 20 px. 各算法在 10 组视频序列上的 DP,如表 1 所示. 由表 1 可知:文中算法在平均 DP 上取得了最优值. 相较于 KCF 算法,平均 DP 提高了 8.2%.

综合以上定性和定量分析可知:文中算法能较好地适应目标尺度及旋转变化,对目标遮挡、光照变

表 1 中心位置误差

Tab. 1 Centre location error

参数	DFT	STC	KCF	文中
CLE/px	56.33	86.30	35.10	26.71
DP/%	37.2	52.4	74.4	82.6

化等也具有一定的鲁棒性;另一方面,通过利用快速傅里叶变换将时域运算变换到频域运算,提高了跟踪效率,具有较好的实时性.

5 结束语

在 KCF 算法的基础上,提出一种利用特征点匹配的方式对目标尺度和角度进行估计的机制.通过与原 KCF 算法和其他当前主流跟踪算法对比,验证了文中算法对尺度和角度估计的有效性.然而,文中算法仍有不足之处.在实验中发现,当目标具有较大角度旋转时,该算法效果不是很好.同时,因为文中算法是根据特征点匹配估计尺度变化和旋转角度变化,当目标区域特征点较少或目标区域较小时,跟踪效果有限.因此,如何处理目标的较大角度变化问题和特征点较少问题将是下一步的工作.

参考文献:

- [1] BOLME D S, BEVERIDGE J R, DRAPER B A. Visual object tracking using adaptive correlation filters[C]// Proceedings of the International Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE Press, 2010: 2544-2550.
- [2] HENRIQUES J F, CASERIO R, MARTINS P. Exploiting the circulant structure of tracking-by-detection with kernels[C]// Proceedings of the Europe Conference on Computer Vision. Florence: Springer Press, 2012: 702-715.
- [3] ZHANG Kaihua, ZHANG Lei, LIU Qingshan, *et al.* Fast visual tracking via dense spatio-temporal context learning [C]// Proceedings of the Europe Conference on Computer Vision. Zurich: Springer Press, 2014: 127-141.
- [4] HENRIQUES J F, CASERIO R, MARTINS P, *et al.* High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(3): 583-596.
- [5] SEVILL L, MILLER E L. Distribution fields for tracking[C]// Proceedings of the International Conference on Computer Vision and Pattern Recognition. Providence: IEEE Press, 2012: 1910-1917.
- [6] ZHANG Kaihua, ZHANG Lei, YANG Minghusan. Real-time compressive tracking[C]// Proceedings of the Europe Conference on Computer Vision. Florence: Springer Press, 2012: 864-877.
- [7] WU Yi, LIM J, YANG M H. Online object tracking: A benchmark[C]// Proceedings of the International Conference on Computer Vision and Pattern Recognition. Portland: IEEE Press, 2013: 2411-2418.
- [8] DANELJAN M, HAGER G, KHAN F S. Accurate scale estimation for robust visual tracking[C]// Proceedings of British Machine Vision Conference. Nottingham: Springer Press, 2014: 1-11.
- [9] 钟必能, 陈雁, 沈映菊. 在线机器学习跟踪算法的进展[J]. 华侨大学学报(自然科学版), 2014, 35(1): 41-46.
- [10] PATNAIK R, CASASENT D. Fast fft-based distortion-invariant kernel filters for general object recognition[C]// Proceedings of the International Society for Optics and Photonics. San Jose: IEEE Press, 2009: 725202.
- [11] ROSTEN E, PORTER R, DRUMMOND T. Faster and better: A machine learning approach to corner detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(3): 105-119.
- [12] CALONDER M, LEPETIT V, STRECHA C, *et al.* BRIEF: Binary robust independent elementary features[C]// Proceedings of the Europe Conference on Computer Vision. Heraklion: Springer Press, 2010: 778-792.
- [13] SHI Jianbo, TOMASI C. Good feature to track[C]// Proceedings of the International Conference on Computer Vision and Pattern Recognition. Seattle: IEEE Press, 1994: 21-23.

(责任编辑: 黄晓楠 英文审校: 吴逢铁)