

# 融合相关滤波与关键点匹配的跟踪算法

张哲\*, 孙瑾, 杨刘涛

南京航空航天大学民航学院, 江苏 南京 211106

**摘要** 提出一种融合相关滤波与关键点匹配的跟踪算法。利用多个基于支持向量机的相关滤波器, 分别对目标进行跟踪和验证, 同时建立并实时更新一个目标和背景关键点数据库。在验证跟踪失败后, 利用关键点匹配的方法对全局关键点进行分类, 根据分类结果对目标关键点进行分析, 从而得到重检测结果。实验结果表明, 在运动模糊、变形、目标遮挡、消失等复杂跟踪场景下, 所提算法比现有算法具有更好的准确性和稳健性。

**关键词** 机器视觉; 目标跟踪; 相关滤波; 关键点匹配; 长期跟踪

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/AOS201939.0215001

## Tracking Algorithm Based on Correlation Filter Fusing with Keypoint Matching

Zhang Zhe\*, Sun Jin, Yang Liutao

College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China

**Abstract** A tracking algorithm is proposed based on correlation filter fusing with keypoint matching. The object is tracked and verified by multiple support-vector-based correlation filters, respectively. Meanwhile, a database containing the keypoints of the target and the background is established and updated in real-time. After the validation of a tracking failure, the global keypoints are classified by utilizing the keypoint matching method and the target keypoints are analyzed according to these classified results. Thus the redetection results are obtained. The experimental results show that the proposed method has a better accuracy and robustness than those by the existing methods in the complex tracking scenes of motion blur, deformation, object occlusion, disappearance and so on.

**Key words** machine vision; object tracking; correlation filter; keypoint matching; long-term tracking

**OCIS codes** 150.1135; 100.4999; 100.3008

## 1 引 言

目标跟踪是计算机视觉领域的重要研究方向之一, 被广泛应用于智能监控、人机交互、自动驾驶导航、行为分析等领域。近年来, 基于视觉的目标跟踪技术有了长足发展, 但是受限于光照变换、目标形变、物体遮挡、快速运动等多种因素, 长期稳定的目标跟踪仍是一个极具挑战性的问题。

目标跟踪技术主要可分为生成式和判别式两类。生成式跟踪方法首先对跟踪目标进行特征建模, 再搜索与模型最相似的目标区域, 并将其作为跟踪预测结果。经典的算法有粒子滤波、卡尔曼滤波、均值漂移<sup>[1]</sup>等。与生成式跟踪方法不同, 判别式跟踪方法将目标跟踪转化为分类问题, 利用学习好的

分类器将跟踪目标从背景中区分出来, 从而得到跟踪结果。判别式跟踪方法由于学习了背景信息, 具有更好的分辨性能, 在识别准确率上普遍优于生成式跟踪方法。判别式相关滤波跟踪算法一经提出, 便以其快速、稳定的跟踪表现, 迅速成为目标跟踪领域的研究热点之一。2010年, Bolme等<sup>[2]</sup>提出了一种最小均方误差和(MOSSE)跟踪器滤波算法, 将相关滤波引入目标跟踪领域。Heriques等<sup>[3]</sup>提出核相关滤波器(KCF), 将梯度方向直方图(HOG)特征与核方法引入相关滤波, 通过将HOG特征映射到非线性空间, 增强特征和滤波器的稳健性。针对跟踪过程中的目标尺度变化等影响, 文献[4-5]在平移滤波的同时增加了尺度滤波, 有效减小了由目标尺度变化引起的模型漂移等的影响。文献[6-7]提出

收稿日期: 2018-07-24; 修回日期: 2018-09-02; 录用日期: 2018-09-17

基金项目: 国家自然科学基金青年科学基金(61702260)、南京航空航天大学研究生创新基地(实验室)开放基金(kfj20170716)

\* E-mail: zhangzhe@nuaa.edu.cn

基于分块的相关滤波跟踪模型,有效提高了目标部分遮挡时的跟踪效果。Zuo 等<sup>[8]</sup>提出的支持相关滤波能有效增大正负样本之间的特征间隔,从而改善滤波器的分类效果。虽然基于相关滤波的跟踪算法在速度和效果上优于生成式跟踪算法,但是在严重遮挡甚至目标消失的情况下,目标跟踪失败后无法进行后续的长期跟踪。

在长期跟踪过程中,目标的跟踪更容易受到多种复杂条件共同作用的干扰,如光照变化、目标遮挡、消失和形变等。这种情况下容易发生目标跟踪失败,从而在跟踪模型中累积误差,最终导致模型漂移直至无法正确跟踪。因此,对于长期跟踪任务,跟踪失败后的目标恢复机制(重检测)不可或缺。Kalal 等<sup>[9]</sup>提出了一种基于经典“训练-学习-检测”机制的长期跟踪算法,其检测器集成多个随机蕨得到的识别结果,与跟踪器得到的目标进行对比验证,得到最终的跟踪结果并更新检测器。但随机蕨应用随机局部二值特征对目标进行编码时,容易受光照等因素的影响。Supančić III 等<sup>[10]</sup>利用自步学习方法结合支持向量机(SVM)筛选出“准确”跟踪的视频帧,并以此帧中的目标对重检测器进行更新。跟踪过程中需要连续更新 SVM,计算量大,导致算法不能实时进行。EBT(EdgeBox Tracking)算法<sup>[11]</sup>根据先验条件对每一帧提取的类物体区域进行排名,对排名靠前的区域再进行检测打分,得到跟踪结果。EBT 算法的全局搜索机制能降低误检率,但是在相相似度高的背景下,跟踪效果差且不能实时进行。相关滤波方法出现后,基于相关滤波的跟踪或重检测的方法也被引入长期跟踪领域。Zhu 等<sup>[12]</sup>在相关滤波的基础上引入一种低秩矩阵近似方法,通过随机采样构建低秩目标表征滤波器,与全局图像进行滤波,用得到的结果对滤波跟踪进行验证与重检测。低秩滤波器只包含对目标表征的重建,缺少相应的背景信息,影响样本的分类。Ma 等<sup>[13]</sup>提出长期相关滤波跟踪(LCT)算法,在相关滤波器外通过更新一个 SVM 分类器对目标进行重检测。但是,较少的更新迭代次数影响了分类器的精度,同时分类器使用的直方图特征缺少目标的局部特征,容易产生误识别。为了更有针对性地进行目标重检测,Wang 等<sup>[14]</sup>提出了高置信度检测更新条件和多峰目标检测方法,从而有效地提高检测效率。跟踪与检测使用同一个模型,虽然能提升跟踪速度,但是无法应对模型漂移问题。

针对以上分析,在此提出一种融合相关滤波和

关键点匹配(CFMK)的长期跟踪框架。首先利用基于 SVM 改进的相关滤波算法<sup>[8]</sup>增强跟踪滤波器的可分辨性,同时训练多个滤波器,融合多滤波响应结果,从而减小噪声误差并对跟踪结果进行验证;其次,针对重检测问题,提出一种基于关键点匹配的目标检测器,通过连续更新目标关键点的时空信息,增强重检测器的稳健性,从而提高整体跟踪框架的准确性和稳定性。

## 2 支持相关滤波跟踪器

相关滤波源于信号处理领域,后被用于图像分类等方面。Bolme 等<sup>[2]</sup>提出的 MOSSE 滤波器对每一帧目标表观进行编码,从而实现快速跟踪,第一次将相关滤波引入目标跟踪领域。在此基础上,Heriques 等<sup>[3]</sup>利用循环矩阵的性质,提出 KCF,将核相关滤波方法和多维 HOG 特征引入目标跟踪,将线性空间滤波扩展至非线性空间,提高了跟踪稳健性。

在 KCF 中,给定以跟踪目标为中心的包含目标及其相应背景的图像块  $\mathbf{x}$ ,其大小为  $M \times N$  个像素。通过  $\mathbf{x}$  分别沿图像行与列方向循环位移,得到密集采样样本  $\mathbf{x}_{m,n}$ ,位移坐标  $(m,n) \in \{0,1,2,\dots,M-1\} \times \{0,1,2,\dots,N-1\}$ ,且满足高斯分布的样本标签  $\mathbf{y}_{m,n}$ 。相关滤波器  $\mathbf{w}$  满足条件:

$$\mathbf{w} = \arg \min_{\mathbf{w}} \sum_{m,n} |\langle \phi(\mathbf{x}_{m,n}), \mathbf{w} \rangle - \mathbf{y}_{m,n}|^2 + \lambda \|\mathbf{w}\|^2, \quad (1)$$

式中: $\phi(\cdot)$  为非线性核空间映射; $\lambda$  为正则项系数; $\langle \cdot, \cdot \rangle$  为向量内积。令  $\mathbf{w} = \sum_{m,n} \alpha_{m,n} \cdot \phi(\mathbf{x}_{m,n})$ ,根据 Heriques 等<sup>[3]</sup>提出的非线性空间方法,将  $\mathbf{w}$  的求解转化为求对偶空间的系数  $\alpha$  ( $\alpha_{m,n}$  为  $\alpha$  的元素)。利用最小二乘法求得闭合解为

$$\alpha = \mathcal{F}^{-1} \left[ \frac{\hat{\mathbf{y}}}{\hat{\phi}^*(\mathbf{x}) \circ \hat{\phi}(\mathbf{x}) + \lambda} \right], \quad (2)$$

式中: $\hat{\cdot}$  和  $\mathcal{F}^{-1}$  分别表示傅里叶变换和傅里叶逆变换; $*$  表示取复共轭; $\circ$  代表 Hadamard 乘积。在下一帧中,用得到的滤波模板的  $\alpha$  和表观模型的  $\mathbf{x}$  对相同大小的搜索窗口样本  $\mathbf{z}$  进行相关滤波操作,得到目标滤波响应  $\bar{\mathbf{y}}$ ,取其最大值作为目标预测结果。 $\bar{\mathbf{y}}$  可表示为

$$\bar{\mathbf{y}} = \mathcal{F}^{-1} [\hat{\alpha} \circ \hat{\phi}^*(\mathbf{x}) \circ \hat{\phi}(\mathbf{z})]. \quad (3)$$

基于 SVM 的支持相关滤波方法<sup>[8]</sup>引入函数间隔,增强正负样本的可分辨性,该方法训练一个支持

相关滤波器  $w$  以及一个偏差  $b$ , 每个目标的滤波响应为

$$\bar{y} = w^T x + b, \quad (4)$$

式中: 上标  $T$  表示共轭转置。根据响应函数和 SVM 函数间隔模型, 训练的  $w$ 、 $b$ 、循环采样样本  $x = [x_1; x_2; \dots; x_{mn}]$  以及对应的样本标签  $y = [y_1; y_2; \dots; y_{mn}]$  应满足目标函数:

$$\begin{cases} \min_{w,b} \|w\|^2 + C \sum_j \xi_j^2, \\ \text{s.t. } y \circ [\mathcal{F}^{-1}(\hat{x}^* \circ \hat{w}) + b\mathbf{1}] \geq 1 - \xi \end{cases}, \quad (5)$$

式中:  $\xi = [\xi_1; \xi_2; \dots; \xi_{mn}]$  为松弛变量;  $C$  为正则系数;  $\mathbf{1}$  表示元素全为 1 的向量;  $j$  代表采样样本序号。根据 Zuo 等<sup>[8]</sup> 的实验设计以及样本中心点与目标中心点的距离设置二分类标签。

设变量  $u = y \circ [\mathcal{F}^{-1}(\hat{x}^* \circ \hat{w}) + b\mathbf{1}] - 1 + \xi$ , 将约束条件代入目标函数得

$$\begin{cases} \min_{w,b,u} \|w\|^2 + C \|y \circ [\mathcal{F}^{-1}(\hat{x}^* \circ \hat{w}) + b\mathbf{1}] - 1 - u\|^2 \\ \text{s.t. } u \geq 0 \end{cases}. \quad (6)$$

针对上述多变量的优化问题, 采取迭代的交替优化算法求  $w$ 、 $b$  的近似解, 每次迭代分步骤更新所求参数:

1) 已知  $w$ 、 $b$ , 更新  $u$ 。由目标函数及约束条件可求得

$$\begin{cases} u = y \circ [\mathcal{F}^{-1}(\hat{x}^* \circ \hat{w}) + b\mathbf{1}] - 1 \\ u = \max(u, 0) \end{cases}. \quad (7)$$

2) 已知  $u$ , 更新  $w$ 、 $b$ 。设  $p = y \circ (1 + u)$ , 根据相关滤波的求解公式, 可得

$$\begin{cases} \hat{w} = \frac{\hat{x} \circ (\hat{p} - b\hat{\mathbf{1}})}{\hat{x}^* \circ \hat{x} + 1/C} \\ b = \mathbf{1}^T [p - \mathcal{F}^{-1}(\hat{x}^* \circ \hat{w})] = \text{mean}(p) \end{cases}. \quad (8)$$

同理, 根据(2)式将样本和滤波器映射到非线性空间中, 迭代更新对偶系数:

$$\hat{a} = \frac{\hat{p} - b\hat{\mathbf{1}}}{\hat{x}^* \circ \hat{x} + 1/C}. \quad (9)$$

### 3 CFMK 算法

CFMK 算法是一种融合支持相关滤波方法和关键点匹配方法的长期目标跟踪算法, 主要由两部分组成: 1) 基于支持相关滤波的跟踪器; 2) 基于关键点匹配的重检测器。整体跟踪流程如图 1 所示。

根据训练好的运动模型对输入帧搜索窗口(图 1 中圆点矩形)进行相关滤波操作, 根据滤波响应值得到预测目标(实线矩形)并利用目标表观模型对预测结果进行可信度检测。对可信度[即平均峰值相关能量值(APCE)] $R_{APCE}$  高于阈值  $\delta_{APCE}$  的跟踪结果]进行尺度估计和模型的更新(图 1 中点状)。对于可信度低于阈值的跟踪结果, 激活目标检测器对目标进行重检测(图 1 中虚线)。利用关键点匹配的方法对全局关键点进行匹配分类, 根据目标关键点得到重检测结果(虚线矩形), 并将该结果与跟踪结果进行比较, 得到最优跟踪结果, 再进行尺度估计和模型的更新。

#### 3.1 基于多滤波器的目标跟踪与验证

为方便比较, 所提算法与主流算法统一采用 HOG 与颜色名<sup>[15]</sup> (CN) 特征描述所有样本, 根据

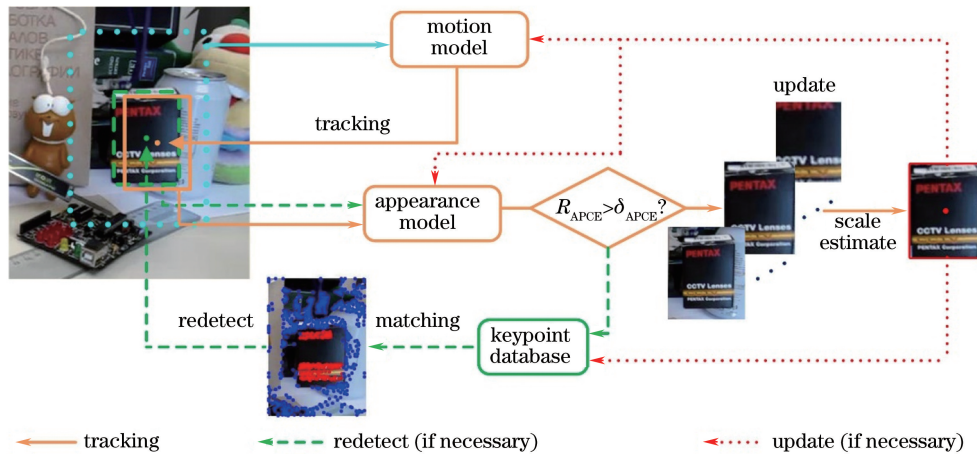


图 1 CFMK 的长期跟踪框架  
Fig. 1 Framework of long-term tracking with CFMK



(7)~(9)式分别对两种特征单独训练支持相关滤波器  $\{\hat{\mathbf{a}}_{\text{HOG}}, \hat{\mathbf{x}}_{\text{HOG}}, b_{\text{HOG}}\}$  和  $\{\hat{\mathbf{a}}_{\text{CN}}, \hat{\mathbf{x}}_{\text{CN}}, b_{\text{CN}}\}$ 。两个滤波器统称运动模型  $\{\hat{\mathbf{a}}_m, \hat{\mathbf{x}}_m, b_m\}$ 。检测时,根据(4)式得到两个运动滤波响应向量,记作  $\bar{\mathbf{y}}_{\text{HOG}}, \bar{\mathbf{y}}_{\text{CN}}$ 。将滤波响应看作目标中心点分布概率,利用 KL (Kullback-Leibler) 散度<sup>[16]</sup>计算优化概率分布  $\bar{\mathbf{y}}_m$ :

$$\begin{cases} \bar{\mathbf{y}}_m = \arg \max \sum_{l=(\text{HOG}, \text{CN})} \sum \mathbf{y}_{m,n}^l \ln \frac{\mathbf{y}_{m,n}^l}{\bar{\mathbf{y}}_{m,n}} \\ \text{s.t.} \quad \sum_{m,n} \bar{\mathbf{y}}_{m,n} = 1 \end{cases}, \quad (10)$$

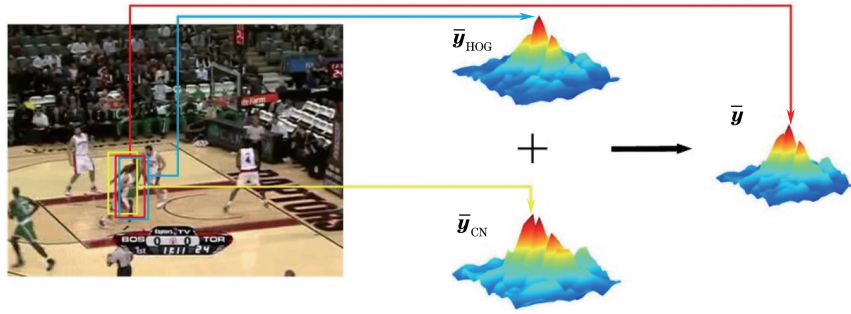


图 2 多特征滤波响应融合

Fig. 2 Response fusion of multiple features

并根据响应计算跟踪结果的可信度。当可信度小于阈值时,激活检测器,对全局进行检索得到新的重检测结果,并将其与跟踪结果进行比较,得到最优预测。所提算法采用文献[14]提出的 APCE 条件计算跟踪可信度:

$$R_{\text{APCE}} = \frac{|\bar{\mathbf{y}}_{\max} - \bar{\mathbf{y}}_{\min}|^2}{\text{mean} \left[ \sum_{m,n} (\bar{\mathbf{y}}_{m,n} - \bar{\mathbf{y}}_{\min})^2 \right]}, \quad (12)$$

式中:  $\bar{\mathbf{y}}_{\max}, \bar{\mathbf{y}}_{\min}$  分别为得到的表观模型滤波响应  $\bar{\mathbf{y}}_a$  的最大值和最小值。该条件能很好地反映滤波响应的波动程度和可信度水平。  $R_{\text{APCE}}$  值越大,目标响应峰值相对背景响应越大,并且除峰值区域外响应更为平滑,目标响应可信度更高,反之响应可信度低。实验中设定阈值  $\delta_{\text{APCE}}$ ,当滤波跟踪结果大于阈值时,更新表观滤波模型与检测器,否则激活目标检测器,对目标进行全局检测。针对目标的尺度变化问题,采用文献[4]的方法在预测位置处构建目标尺度金字塔,通过尺度滤波得到最佳的尺度变化率。

运动模型参数每帧都会更新,而表观模型和重检测器只在跟踪目标可信度高于阈值时进行更新。为减小运动模糊、噪声干扰和遮挡等条件的影响,采用线性插值法更新模型:

式中:  $l$  为特征类型;  $m, n$  为二维坐标值;  $\bar{\mathbf{y}}_{m,n}$  为对应位置元素值。利用拉格朗日乘子法求解(10)式,得到

$$\bar{\mathbf{y}}_m = \frac{\bar{\mathbf{y}}_{\text{HOG}} + \bar{\mathbf{y}}_{\text{CN}}}{2}, \quad (11)$$

取最大值作为目标中心点预测,如图 2 所示。

为了判断是否出现跟踪失败,除了训练区分目标和背景的滤波器(运动模型)之外,还额外训练一个只包含目标的表观滤波模型  $\{\hat{\mathbf{a}}_a, \hat{\mathbf{x}}_a, b_a\}$ ,其目的在于计算跟踪结果与目标表观的滤波响应向量  $\bar{\mathbf{y}}_a$ ,

$$\begin{cases} \hat{\mathbf{x}}_m^{(t)} = (1 - \gamma)\hat{\mathbf{x}}_m^{(t-1)} + \hat{\mathbf{x}}_m \\ \hat{\mathbf{a}}_m^{(t)} = (1 - \gamma)\hat{\mathbf{a}}_m^{(t-1)} + \hat{\mathbf{a}}_m \\ b_m^{(t)} = (1 - \gamma)b_m^{(t-1)} + b_m \end{cases}, \quad (13)$$

式中:  $\gamma$  为参数学习率;  $t$  表示第  $t$  帧。

### 3.2 基于关键点匹配的目标检测器

在支持相关滤波跟踪器的基础上,针对目标消失和严重遮挡的情况,提出一种基于关键点匹配的目标检测器。检测器通过融合多帧图像中目标与背景关键点的时空信息建立数据库,在重检测时利用关键点匹配的原则对当前帧关键点进行分类,再对分类结果进行分析,从而得到重检测结果。

关键点数据库  $D = T \cup B$  包含目标关键点集合  $T$  与背景关键点集合  $B$ :

$$T = \{(d_i, r_i)\}_{i=1}^{N_T}, B = \{d_i\}_{i=1}^{N_B}. \quad (14)$$

关键点选择具有良好尺度和旋转不变性的 BRISK<sup>[17]</sup> (Binary Robust In variant Scalable Keypoints)特征,其中:  $d_i \in \mathbf{R}^{512}$  表示关键点二进制特征描述子,特征维度为 512;  $r_i$  表示目标关键点与目标中心点的相对位置;  $N_T, N_B$  为关键点数量。由于剧烈运动或者物体形变,目标表观特征也随之改变。为了更好地适应特征的变化,不同于文献[18]利用第一帧初始化关键点集合的方式,提出一种融合多帧图像关键点并实时更新的关键点数据库,降



低由目标表观变化带来的影响。设目标关键点集合  $T$  由  $K$  帧图像中目标关键点子集  $V$  构成:

$$T = V_{t-1} \cup V_{t-2} \cup \dots \cup V_{t-K}, \quad (15)$$

式中:  $t$  表示当前帧。只有在可置信度满足阈值条件时更新数据库,所以子集可能来自非连续帧。更新时,根据关键点时效性,用当前帧关键点子集  $V_t$  替换时域最远帧的子集  $V_{t-K}$ 。更新关键点与原有关键点存在的冗余会导致重检测匹配时的干扰和额外的时间消耗,在更新时先对  $V_t$  和  $T - V_{t-K}$  进行一次匹配,过滤  $T - V_{t-K}$  中的重复关键点,加入目标关键点集合:

$$\begin{cases} V_m = \text{match}(V_t, T - V_{t-K}) \\ T_{\text{new}} = V_t + T - V_{t-K} - V_m \end{cases}, \quad (16)$$

式中:  $\text{match}()$  为匹配函数。在需要重检测的视频帧中的全局图像中提取候选关键点  $P = \{(d_i, a_i)\}_{N_{p_i}=1}$ , 其中  $a_i$  表示关键点在视频帧中的绝对位置,  $N_{p_i}$  为候选关键点数量。在进行关键点匹配时,计算  $P$  中关键点  $p_i$  与数据库  $D$  中每个关键点的 Hamming 距离,并采用最近邻方法得到最近的两个关键点  $d_i^{1\text{st}}, d_i^{2\text{nd}}$ , 距离分别为  $h^{1\text{st}}$  和  $h^{2\text{nd}}$ ,  $P$  中关键点的匹配分类准则为

$$\begin{cases} p_i \in T, & \text{if } d_i^{1\text{st}} \in T \ \& \ h_i^{1\text{st}} < \theta_c \ \& \ h_i^{1\text{st}}/h_i^{2\text{nd}} < \theta_r \\ p_i \in B, & \text{otherwise} \end{cases}, \quad (17)$$

式中:  $\theta_c, \theta_r$  分别为置信阈值和比例阈值;  $P$  中分类为目标关键点的  $p_i \in T$ , 加上与其匹配的关键点相对位置  $r_i$ , 组成子集  $P^T = \{(d_i, a_i, r_i)\}$ 。

根据匹配得到的目标关键点子集  $P^T$  的位置信息, 计算每个点对目标中心点的预测:

$$a_i^{\text{center}} = a_i - r_i. \quad (18)$$

考虑到目标的尺度变换, 关键点与中心点的相对距离也会随之改变。根据跟踪时的尺度变换系数  $s$  对(18)式进行修正, 可得

$$a_i^{\text{center}} = a_i - s \cdot r_i. \quad (19)$$

$P^T$  中所有关键点对应的目标中心点预测构成集合  $M$ 。匹配过程中可能出现错误且位置计算中可能产生误差, 导致预测中心点位置可能在目标轮廓外。为了避免少数错误预测的干扰, 采用自下而上的层次聚类法, 以点之间的欧氏距离为相似度测量, 对  $M$  中所有预测中心点进行聚类, 得到子集  $Q_1, Q_2, \dots, Q_n$ 。子集数量  $n$  由截断阈值  $\sigma$  决定, 最终选择元素数量最多的子集对中心点进行预测, 并取子集中所有元素的均值作为重检测器的检测结果。

## 4 实验及结果分析

### 4.1 实验参数与数据

实验环境为 Intel i5-4590@3.30 GHz, 算法开发采用 Matlab 与 OpenCV 库接口。实验参数设置为:  $C=10^4$ , 样本标签分类的上下限阈值  $\delta_u=0.3$ ,  $\delta_l=0.6$ ; 置信度函数尺度参数  $\lambda=0.5$ , 形状参数  $\beta=2$ ; 模型更新时的参数学习率  $\gamma=0.025$ ; 跟踪可置信度阈值  $\delta_{\text{APCE}}=10$ ; 层次聚类截断阈值  $\sigma=4$ 。

选择公开数据库 OTB-13<sup>[19]</sup> 和 OTB-15<sup>[20]</sup> 进行实验。OTB-13 数据库中包含 50 段视频序列, 可以分为 11 种不同的视觉跟踪挑战场景属性, 如尺度变化、目标遮挡、光照变化、运动模糊、物体形变等, 每段视频序列可能有多个属性。OTB-15 数据库是在 OTB-13 数据库的基础上增加 50 组测试视频序列得到的。增加的序列场景属性更为复杂, 跟踪难度更大。

选择评价准则 OPE (One-Pass Evaluation) 计算跟踪准确率和成功率, 以评估跟踪性能。准确率为跟踪中心点位置误差小于阈值的帧数  $F_{\text{center}}$  占总体视频帧数  $F_{\text{total}}$  的比例, 即  $F_{\text{center}}/F_{\text{total}}$ , 成功率则表示跟踪重叠率(交并比)大于阈值的帧数  $F_{\text{overlap}}$  占比, 即  $F_{\text{overlap}}/F_{\text{total}}$ 。

### 4.2 连续 $K$ 帧关键点影响分析

目标关键点数据集  $T$  对目标重检测结果有着直接影响,  $T$  是由连续  $K$  帧目标关键点子集组成的, 因此  $K$  的取值对跟踪结果有一定影响。通过实验比较不同  $K$  值下跟踪算法的平均距离精度(DP)和平均重叠率精度(OP), 即不同位置误差阈值和重叠率阈值下平均跟踪准确率和成功率, 实验结果如表 1 所示。

由表 1 可知, 数据库为 OTB-13 时,  $K=1$  时跟踪准确率和跟踪精度都更高, 当数据库扩展为表 1 不同  $K$  值下所提算法在不同数据库上的平均 DP 和平均 OP

Table 1 Mean DP and OP on different databases for proposed algorithm under different values of  $K$

Value of $K$	Mean DP		Mean OP	
	OTB-13	OTB-15	OTB-13	OTB-15
1	0.885	0.796	0.642	0.569
2	0.871	0.788	0.630	0.562
3	0.874	0.796	0.633	0.569
4	0.871	0.795	0.632	0.568
5	0.871	0.795	0.633	0.567

OTB-15 时,虽然  $K=1$  与  $K=3$  时跟踪准确率相同且均优于其他取值,但是由于  $K=1$  时 OTB-13 数据库效果较好,说明  $K=1$  时 100 个视频帧的平均准确率波动更大,因此数据库为 OTB-15 时,取  $K=3$ 。

### 4.3 跟踪算法性能比较

为验证所提跟踪算法的性能,选取 8 种近年来出现的效果良好的跟踪算法 CSK<sup>[21]</sup> (Circulant Structure of Tracking)、KCF<sup>[3]</sup>、DSST<sup>[4]</sup> (Discriminative Scale Space Tracking)、SKSCF<sup>[8]</sup>

(Scale Kernel Support Correlation Filter)、LCT<sup>[13]</sup>、BACF<sup>[22]</sup> (Background-Aware Correlation Filter)、SRDCF<sup>[23]</sup> (Spatially Regularized Correlation Filters)、STAPLE<sup>[24]</sup> (Sum of Template And Pixel-wise LEarners) 进行对比,这些算法都是基于相关滤波或相关滤波改进的算法。表 2 列出了 9 种跟踪算法在相关数据集上运行的平均每秒传输帧数(FPS)、平均 DP 和平均 OP。图 3 则给出了 9 种算法分别在数据库 OTB-13 和 OTB-15 上相应的跟踪成功率曲线和跟踪准确率曲线。

表 2 9 种跟踪算法在不同数据库上的平均跟踪性能参数

Table 2 Mean tracking performance parameters on different databases for nine kinds of tracking algorithms

Algorithm	Mean FPS (OTB-15) / (frame · s <sup>-1</sup> )	Mean DP		Mean OP	
		OTB-13	OTB-15	OTB-13	OTB-15
CSK	440.41	0.545	0.518	0.398	0.382
KCF	98.49	0.704	0.675	0.500	0.470
DSST	46.44	0.737	0.693	0.554	0.520
LCT	24.18	0.848	0.762	0.593	0.527
SKSCF	39.29	0.864	0.772	0.623	0.549
STAPLE	76.70	0.782	0.784	0.593	0.578
SRDCF	5.42	0.838	0.789	0.626	0.598
BACF	32.12	0.849	0.817	0.645	0.616
Proposed	29.83	0.885	0.796	0.642	0.569

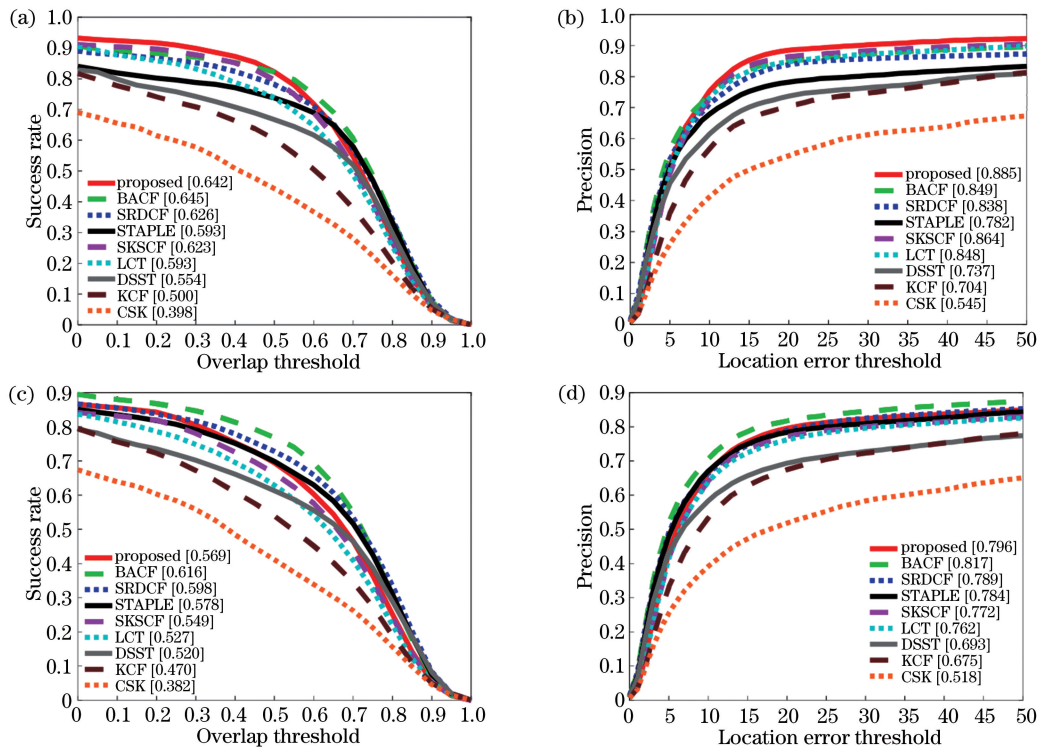


图 3 9 种算法在不同数据库上的跟踪结果。OTB-13 数据库上的(a)跟踪成功率和(b)跟踪准确率;OTB-15 数据库上的(c)跟踪成功率和(d)跟踪准确率

Fig. 3 Tracking results for nine kinds of trackers on different databases. (a) Tracking success rate and (b) tracking precision for OTB-13 database; (c) tracking success rate and (d) tracking precision for OTB-15 database

结合表 2 数据与图 3 中曲线可知,在 OTB-13 数据库上,所提算法的跟踪成功率接近 BACF 算法,且高于其他算法。所提算法的跟踪精度也优于其他 8 种算法。在 OTB-15 数据库上,所提算法的跟踪成功率接近 STAPLE 算法,低于针对边界效应问题的 BACF、SRDCF 算法,但是在跟踪精度上仅次于 BACF 算法而优于其他跟踪算法。由表 2 可知,在统一的实验环境下,所提算法在平均跟踪速度

上能够达到  $29.83 \text{ frame} \cdot \text{s}^{-1}$ ,与 BACF 算法相近,明显高于 SRDCF 算法,能基本满足跟踪实时性要求。

为充分对比各个算法针对不同场景的性能差异,将 OTB-13 数据库中的视频序列按照场景属性分别进行跟踪成功率比较,结果如图 4 所示。从图 4 可知,所提的长期跟踪算法在 OTB-13 数据库的大部分场景属性中均有良好表现,尤其是在目标遮挡

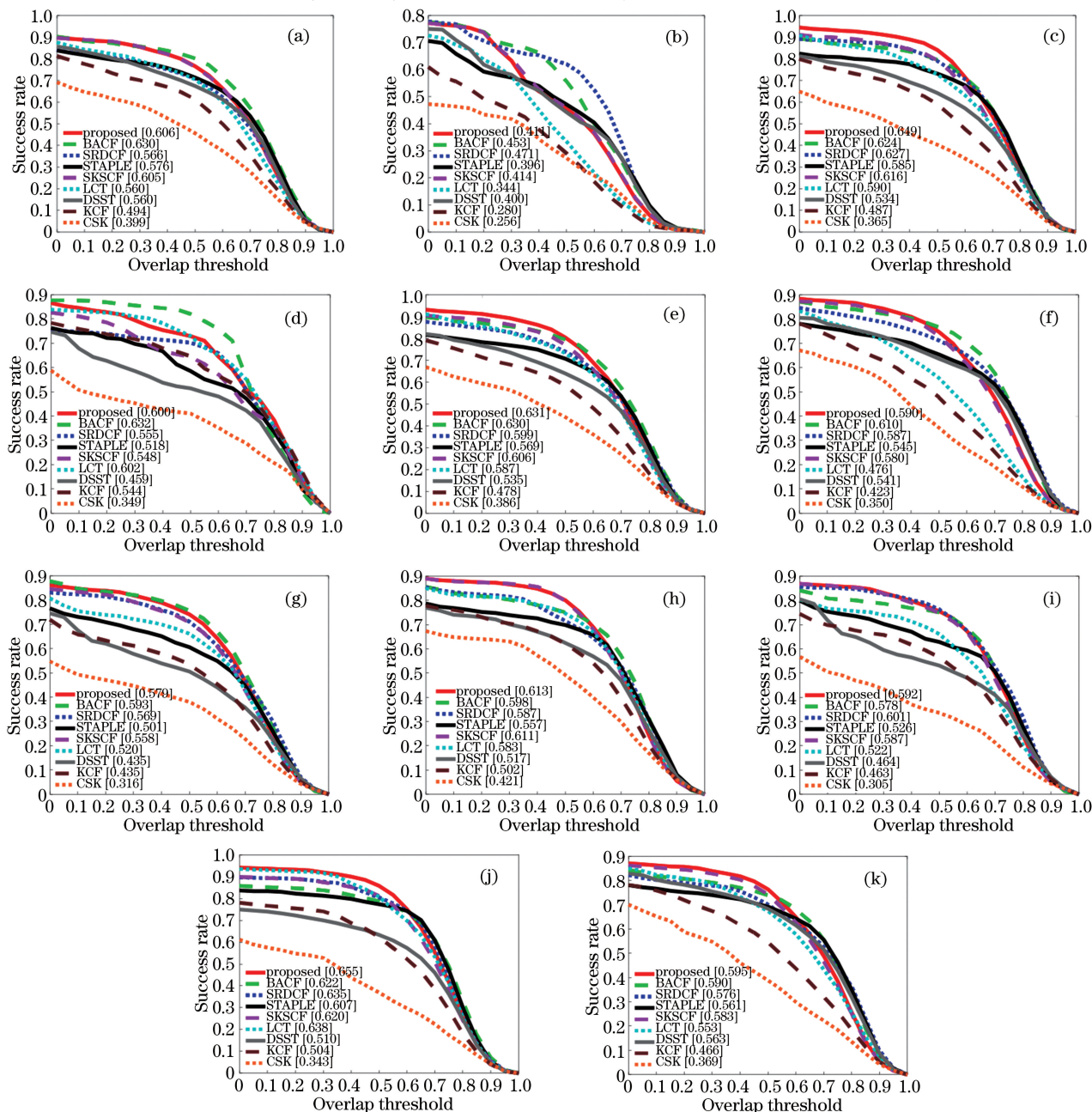


图 4 11 种场景下各算法的跟踪成功率比较。(a)平面内旋转;(b)低分辨率;(c)目标遮挡;(d)出视野;(e)平面外旋转;(f)尺度变换;(g)快速运动;(h)背景杂乱;(i)运动模糊;(j)物体变形;(k)光照变化

Fig. 4 Comparison of tracking success rate for each algorithm under eleven kinds of scenes. (a) In-plane rotation; (b) low resolution; (c) object occlusion; (d) out of view; (e) out-of-plane rotation; (f) scale variation; (g) fast motion; (h) background clutter; (i) motion blur; (j) object deformation; (k) illumination variation



[图 4(c)]、背景杂乱[图 4(h)]、物体变形[图 4(j)]和光照变化[图 4(k)]等场景中。同时,所提算法的滤波跟踪器基于 SKSCF 算法,是尺度自适应的支持相关滤波算法。可以看出,对跟踪结果进行判断并加入基于关键点匹配的重检测器后,虽然跟踪速度有所下降,但是算法的平均准确率和成功率均提升了约 2%。图 4 则显示尤其针对存在出视野[图 4(d)]、物体变形[图 4(j)]和目标遮挡[图 4(c)]属性的视频场景,较 SKSCF 算法的平均跟踪成功率分别提高了 5.2%、3.5% 和 3.3%。

为更加直观地显示不同算法的跟踪效果,选取了跟踪精度较高的 5 种算法(BACF、SRDCF、STAPLE、SKSCF 和 LCT)与所提算法在 6 组视频序列中的跟踪结果进行对比,如图 5 所示。序列①存在快速运动以及运动模糊现象,实验结果显示经

历多次快速运动和运动模糊(0284、0624 帧)后,只有所提算法和 SRDCF 算法能长期、准确地跟踪到目标;序列②在跟踪过程中存在密集的相似目标干扰和目标旋转的现象,结果显示所提算法能排除环境干扰,持续稳定地跟踪到目标;序列③、④、⑤均存在目标严重遮挡或消失,序列④、⑤还伴随着人体目标形变的影响。结果显示,所提算法在三组序列中都能实现稳定跟踪。其余 5 种算法,只有 BACF、SRDCF 和 LCT 算法能在部分情况下跟踪到目标,显示了所提算法在目标遮挡和目标消失情况下稳定的跟踪性能;序列⑥不仅有目标遮挡、运动模糊、物体形变等属性,还存在着光照变化的影响(0278 帧),在这种光照条件下,所提算法依然能长期、稳定地跟踪到目标,说明了所提算法对光照条件变化的稳健性。

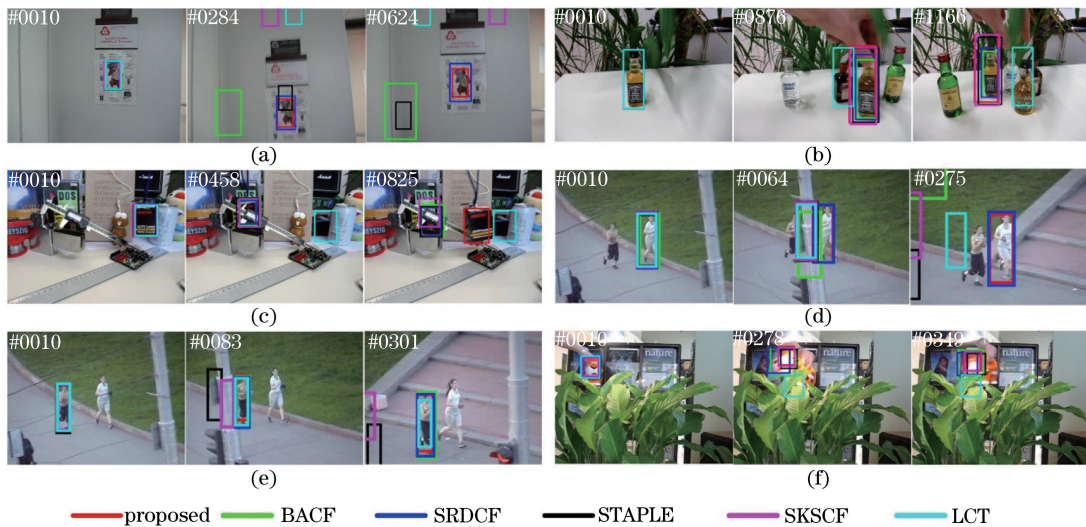


图 5 6 种跟踪算法在部分序列上的跟踪结果图。(a) BlurOwl; (b) Liquor; (c) Box; (d) Jogging 1; (e) Jogging 2; (f) Tigger  
Fig. 5 Tracking result graphs in partial sequences for six kinds of tracking algorithms. (a) BlurOwl; (b) Liquor;  
(c) Box; (d) Jogging 1; (e) Jogging 2; (f) Tigger

## 5 结 论

针对目标遮挡、消失、变形等严重干扰的长期目标跟踪问题,提出一种融合支持相关滤波和关键点匹配的重检测长期跟踪策略。利用支持相关滤波训练多滤波模型,减小随机误差,并对跟踪结果进行验证;同时建立并更新包含目标和背景关键点时空信息的数据库,当判断目标跟踪失败时,根据关键点匹配原则对候选关键点进行分类,对分类结果进行分析得到目标重检测结果。实验结果表明,所提跟踪算法能有效提高目标跟踪的准确性和稳健性,尤其对于目标消失等严重的干扰场景具有很好的性能。所提算法的跟踪速度能基本满足实时性要求,且仍

有可提升空间。但所提算法没有考虑相关滤波模型引起的边界效应,对在视点边缘的目标跟踪有一定影响。下一步工作需要考虑加入边界效应的影响,从而进一步提高跟踪效果。

## 参 考 文 献

- [1] Vojir T, Noskova J, Matas J. Robust scale-adaptive mean-shift for tracking[J]. Pattern Recognition Letters, 2014, 49: 250-258.
- [2] 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, 2010: 2544-2550.
- [3] 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.
- [4] Danelljan M, Häger G, Khan F S, *et al.* Discriminative scale space tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(8): 1561-1575.
- [5] Li Y, Zhu J K. A scale adaptive kernel correlation filter tracker with feature integration[C]. European Conference on Computer Vision, 2014: 254-265.
- [6] Liu T, Wang G, Yang Q X. Real-time part-based visual tracking via adaptive correlation filters[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 4902-4912.
- [7] Liu S, Zhang T Z, Cao X C, *et al.* Structural correlation filter for robust visual tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 4312-4320.
- [8] Zuo W M, Wu X H, Lin L, *et al.* Learning support correlation filters for visual tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018: 2829180.
- [9] Kalal Z, Mikolajczyk K, Matas J. Tracking-learning-detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [10] Supančič III J S, Ramanan D. Self-paced learning for Long-term tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013: 2379-2386.
- [11] Zhu G, Porikli F, Li H D. Beyond local search: tracking objects everywhere with instance-specific proposals[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 943-951.
- [12] Zhu G B, Wang J Q, Wu Y, *et al.* Collaborative correlation tracking[C]. Proceedings of the British Machine Vision Conference, 2015: 184.
- [13] Ma C, Yang X K, Zhang C Y, *et al.* Long-term correlation tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 5388-5396.
- [14] Wang M M, Liu Y, Huang Z Y. Large margin object tracking with circulant feature maps[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 4800-4808.
- [15] Danelljan M, Khan F S, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014: 1090-1097.
- [16] Ma C, Huang J B, Yang X K, *et al.* Adaptive correlation filters with long-term and short-term memory for object tracking[J]. International Journal of Computer Vision, 2018, 126(8): 771-796.
- [17] Leutenegger S, Chli M, Siegwart R Y. BRISK: binary robust invariant scalable keypoints[C]. International Conference on Computer Vision, 2011: 2548-2555.
- [18] Nebel G, Pflugfelder R. Clustering of static-adaptive correspondences for deformable object tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015: 2784-2791.
- [19] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013: 2411-2418.
- [20] Wu Y, Lim J, Yang M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1834-1848.
- [21] 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. Heidelberg: Springer, 2012, 7575: 702-715.
- [22] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking[C]. IEEE International Conference on Computer Vision (ICCV), 2017: 1144-1152.
- [23] Danelljan M, Häger G, Khan F S, *et al.* Learning spatially regularized correlation filters for visual tracking[C]. IEEE International Conference on Computer Vision (ICCV), 2015: 4310-4318.
- [24] Bertinetto L, Valmadre J, Golodetz S, *et al.* Staple: complementary learners for real-time tracking [C]. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 1401-1409.