

引用格式:袁友前,黄山. 结合颜色模型的核相关滤波跟踪算法[J]. 电光与控制,2019,26(11):51-55. YUAN Y Q, HUANG S. Kernelized correlation filter tracking algorithm combined with color model[J]. Electronics Optics & Control, 2019, 26(11):51-55.

## 结合颜色模型的核相关滤波跟踪算法

袁友前, 黄山  
(四川大学,成都 610065)

**摘要:** 为改善 Staple 目标跟踪算法的运行速度,在学习位置滤波器的过程中,对提取的图像特征进行 PCA 降维;在学习尺度滤波器的过程中,将提取的不同尺度样本个数由 33 减少至 17,并且通过 QR 分解对尺度信息进行压缩。为保证尺度估计的准确性,在计算尺度响应时,使用插值法将尺度响应个数插值到 33。实验结果显示,在准确率几乎不变的情况下,所提算法可将跟踪速度提升 50% 左右。

**关键词:** 目标跟踪; PCA 降维; 位置滤波器; 尺度滤波器; QR 分解

**中图分类号:** TP391 **文献标志码:** A **doi:**10.3969/j.issn.1671-637X.2019.11.011

## Kernelized Correlation Filter Tracking Algorithm Combined with Color Model

YUAN You-qian, HUANG Shan  
(Sichuan University, Chengdu 610065, China)

**Abstract:** To improve the speed of Staple target tracking algorithm, dimension reduction of the extracted image features is carried out by using Principal Component Analysis (PCA) in the process of learning translation filters. In the process of learning scale filters, the number of extracted samples with different scales is reduced from 33 to 17, and scale reduction is carried out by using QR-factorization. Scale response number is increased to 33 by interpolation to ensure scale estimation accuracy. Simulation results show that detection accuracy is almost unchanged and the tracking speed is improved by about 50% in comparison with Staple algorithm.

**Key words:** target tracking; PCA dimension reduction; translation filter; scale filter; QR-factorization

### 0 引言

视频目标跟踪<sup>[1]</sup>是机器视觉中的一项重要任务。2010年 DANELLJAN 等首次将核相关滤波应用在了跟踪领域<sup>[2]</sup>,近几年在此基础上产生了许多优异的跟踪算法<sup>[3-7]</sup>。其中,DANELLJAN 等提出了 DSST<sup>[8]</sup>(Discriminative Scale Space Tracking)跟踪算法,使得跟踪器能够确定目标的位置信息和尺度信息,该算法对于运动模糊、光照变化以及背景目标颜色相似的跟踪场景具有鲁棒性,但是在目标快速运动及形变的情况下,跟踪效果较差。而采用颜色特征统计的算法对目标形变以及快速运动并不敏感,如 DAT(Distractor-Aware Tracking)<sup>[9]</sup>。Staple<sup>[10]</sup>算法将这两种算法高效地结合起来,在目标快速形变、背景颜色干扰、运动模糊的场景下都能准确对

目标进行跟踪,该算法对每个像素进行特征提取,具有冗余性,本文算法对此做出改进。

### 1 Staple 算法

#### 1.1 Staple 算法思想

Staple 算法根据响应分数的大小来确定目标的位置和尺寸。其中,目标的位置  $f(x)$  由位置滤波器分数  $f_{\text{templ}}(x)$  和直方图分数  $f_{\text{hist}}(x)$  的线性组合决定,即

$$f(x) = (1 - \omega)f_{\text{templ}}(x) + \omega f_{\text{hist}}(x) \quad (1)$$

式中: $\omega$  是合并因子。

目标的尺寸由尺度滤波器分数决定。

#### 1.2 滤波器分数

根据滤波器的最小二乘公式,每张图片的损失函数为

$$l_{\text{templ}}(x) = \|g - \sum h \cdot f\|^2 + \lambda \sum \|h\|^2 \quad (2)$$

式中: $f$  表示  $d$  维输入样本特征图,对应期望输出为高

收稿日期:2018-10-15

修回日期:2018-12-18

作者简介:袁友前(1995—),男,四川遂宁人,硕士生,研究方向为视频目标跟踪。

斯函数  $g; h$  为对应的滤波器。将式(2)转换到频域, 通过最小化  $l_{\text{tmpl}}(x)$  可得到滤波器  $H$ , 以第  $t-1$  帧为例

$$H_{t-1} = \frac{\bar{G}F}{\sum_{k=1}^d \bar{F}^k F^k + \lambda} = \frac{A_{t-1}}{B_{t-1}} \quad (3)$$

为减少更新滤波器的计算量, 在  $t-1$  帧目标位置周围提取训练样本的 FHOG 特征图  $F_{t-1}$ , 用来更新滤波器的分子和分母, 即

$$\begin{cases} A_{t-1} = (1 - \eta_{\text{tmpl}})A_{t-2} + \eta_{\text{tmpl}} \bar{G} F_{t-1} \\ B_{t-1} = (1 - \eta_{\text{tmpl}})B_{t-2} + \eta_{\text{tmpl}} \bar{F}_{t-1} F_{t-1} \end{cases} \quad (4)$$

式中:  $\eta_{\text{tmpl}}$  为滤波器学习率。

在第  $t$  帧图片  $x_t$  中, 从上一帧目标位置  $p_{t-1}$  周围提取测试样本  $Z_t$ , 与上一帧训练得到的滤波器  $H_{t-1}$  进行卷积, 并进行离散傅里叶逆变换 (IDFT) 得到相应的滤波器分数

$$f_{\text{tmpl}}(x_t) = F^{-1} \left\{ \frac{\sum_{i=1}^d \bar{A}_{t-1} Z_t^i}{B_{t-1} + \lambda} \right\}. \quad (5)$$

当检测到第  $t$  帧目标后, 再根据式(4)对第  $t$  帧滤波器进行更新, 用以对下一帧目标进行跟踪。

### 1.3 直方图分数

在第  $t-1$  帧, 分别对目标区域  $O$  和背景区域  $B$  的像素进行 one-hot 编码, 目标在每个特征维度的损失函数为

$$l_{\text{hist}}(x, \beta) = \sum_{j=1}^M \left[ \frac{N^j(O)}{|O|} \cdot (\beta^j - 1)^2 + \frac{N^j(B)}{|B|} \cdot (\beta^j)^2 \right] \quad (6)$$

式中:  $N^j(*) = |\{u \in *\}|$  表示在区域  $*$  的  $j$  维特征像素数量。令式(6)最小化可得直方图模板

$$\beta_{t-1}^j = \frac{\rho^j(O)}{\rho^j(O) + \rho^j(B) + \lambda} \quad j=1, \dots, M \quad (7)$$

式中:  $\rho^j(*) = \frac{N^j(*)}{|*|}$  表示在  $*$  单位区域中  $j$  维特征像素的分布概率。为方便计算且达到近似鲁棒效果, 采用式(8)对直方图模板进行更新

$$\begin{cases} \rho_{t-1}(O) = (1 - \eta_{\text{hist}})\rho_{t-2}(O) + \eta_{\text{hist}} \rho'_{t-1}(O) \\ \rho_{t-1}(B) = (1 - \eta_{\text{hist}})\rho_{t-2}(B) + \eta_{\text{hist}} \rho'_{t-1}(B) \end{cases} \quad (8)$$

式中:  $\rho_t(*)$  表示由  $j=1, \dots, M$  个  $\rho_t^j(*)$  组成的向量;  $\eta_{\text{hist}}$  为直方图学习率。从第  $t$  帧提取一个特征图像  $\phi_t$ , 其直方图分数为  $f_{\text{hist}}(x) = \beta_{t-1}^T \phi_t$ 。

算法流程和直方图分数如图1所示。

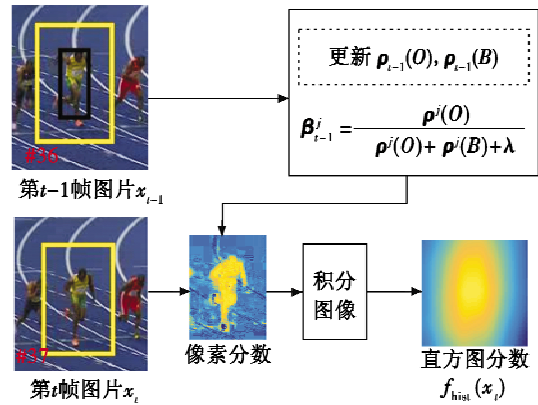


图1 直方图分数

Fig. 1 Histogram scores

## 2 基于 Staple 的改进算法

### 2.1 位置滤波器的改进

在第  $t-1$  帧目标位置周围提取  $N \times D$  维 FHOG 特征样本  $F_{t-1} = (X_1, X_2, \dots, X_D)$ , 其中,  $X_i = (X_i^{(1)}, X_i^{(2)}, \dots, X_i^{(N)})$ 。令每一列数据分别减去该列数据的平均值, 得到  $N \times D$  维去均值矩阵  $F_r = (x_1, x_2, \dots, x_D)$ 。根据协方差公式, 即

$$\text{cov}(X_m, X_n) = \sum_{j=1}^N \frac{X_m^{(j)} X_n^{(j)}}{N-1} \quad (9)$$

构成  $D \times D$  维协方差矩阵  $C$ , 通过求解特征方程

$$(C - \lambda I)U = 0 \quad (10)$$

得到特征值  $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_D\}$  和其对应的特征向量  $U = \{u_1, u_2, \dots, u_D\}$ , 选取最大的前  $\tilde{D}$  个特征值所对应的特征向量组成  $D \times \tilde{D}$  维投影矩阵  $T_{t-1}$ , 通过主成分映射得到降维后的滤波器训练样本  $\tilde{F}_{t-1} = F_r \times T_{t-1}$ 。根据式(4), 第  $t-1$  帧滤波器的分子、分母更新公式变为

$$\begin{cases} \tilde{A}_{t-1} = (1 - \eta_{\text{tmpl}})\tilde{A}_{t-2} + \eta_{\text{tmpl}} \bar{G} \tilde{F}_{t-1} \\ \tilde{B}_{t-1} = (1 - \eta_{\text{tmpl}})\tilde{B}_{t-2} + \eta_{\text{tmpl}} \bar{F}_{t-1} \tilde{F}_{t-1} \end{cases} \quad (11)$$

如图2所示, 由式(11)得到第  $t-1$  帧位置滤波器  $\tilde{H}_{t-1} = \tilde{A}_{t-1} / \tilde{B}_{t-1}$  后, 按上述方法对第  $t$  帧测试样本进行降维得到  $\tilde{Z}_t$ , 与位置滤波器  $\tilde{H}_{t-1}$  卷积并进行 IDFT 得到第  $t$  帧的位置滤波器分数

$$\tilde{f}_{\text{tmpl}}(x_t) = F^{-1} \left\{ \frac{\sum_{i=1}^{\tilde{D}} \tilde{A}_{t-1} \tilde{Z}_t^i}{\tilde{B}_{t-1} + \lambda} \right\}. \quad (12)$$

根据式(1), 由位置滤波器和直方图的线性组合分数最大值来确定目标在第  $t$  帧中的位置  $P_t$ , 如图3所示。

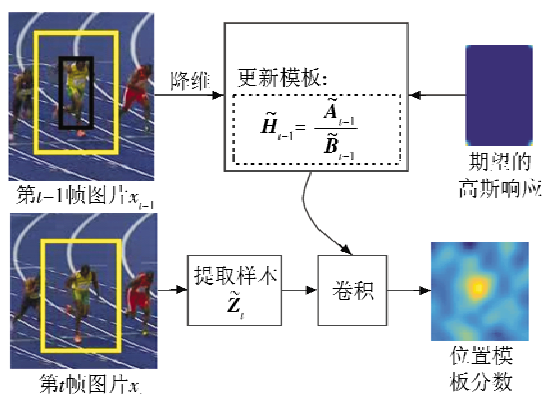


图 2 位置滤波器分数

Fig. 2 Translation filter scores

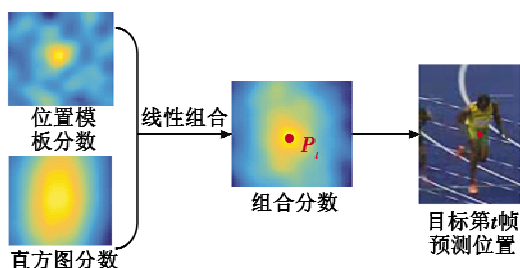


图 3 目标位置

Fig. 3 Target position

### 2.2 尺度滤波器的改进

原 Staple 算法在学习尺度滤波器和进行尺度检测时,提取样本的尺度数量  $S$  为 33,尺度选择方式为

$$\alpha^a P \times \alpha^a E \quad n \in \left\{ \left[ -\frac{s-1}{2}, \dots, \frac{s-1}{2} \right] \right\} \quad (13)$$

式中: $P, E$  分别是目标的宽和高;  $\alpha$  为尺度因子。为减少计算量,本文算法对式(13)中  $n \in \{[-16, \dots, 16]\}$  依次间隔取值从而得到本文算法提取样本的尺度,尺度个数压缩为 17,再使用 QR 分解将每个尺度样本的特征维度压缩到 17。在训练第  $t-1$  帧尺度滤波器时,于目标位置  $P_{t-1}$  周围提取  $K \times \tilde{S}$  维尺度样本  $V_{t-1}$ ,并进行 QR 分解,即

$$V_{t-1} = Q_{t-1} R_{t-1} \quad (14)$$

式中: $Q_{t-1}$  为  $K \times \tilde{S}$  维正交矩阵; $R_{t-1}$  为  $\tilde{S} \times \tilde{S}$  维上三角矩阵。令  $\tilde{V}_{t-1} = Q_{t-1}^T V_{t-1}$ ,得到压缩后的  $\tilde{S} \times \tilde{S}$  维尺度测试样本  $\tilde{V}_{t-1}$ 。由式(4)可得,第  $t-1$  帧尺度滤波器分子、分母更新算式变为

$$\begin{cases} A'_{t-1} = (1 - \eta_{scale}) A'_{t-2} + \eta_{scale} \bar{G} \tilde{V}_{t-1} \\ B'_{t-1} = (1 - \eta_{scale}) B'_{t-2} + \eta_{scale} \tilde{V}_{t-1} \tilde{V}_{t-1}^T \end{cases} \quad (15)$$

式中: $\eta_{scale}$  为尺度滤波器学习率。第  $t-1$  帧尺度滤波器为  $H'_{t-1} = A'_{t-1} / B'_{t-1}$ 。确定第  $t$  帧目标位置后,在  $P_t$  位置周围提取尺度测试样本  $Y_t$ ,按照上述方法对其进行压缩  $\tilde{Y}_t = Q_{t-1}^T Y_t$ ,与上一帧的尺度滤波器  $H'_{t-1}$  进行卷积得到相应的尺度响应值为

$$y_t = \sum_{i=1}^{\tilde{S}} \tilde{A}'_{t-1} \tilde{Y}_t^i / (\tilde{B}'_{t-1} + \lambda) \quad (16)$$

此时只有 17 个尺度响应值,为了对尺度估计更加精确,将响应值补 0,插值为 33 个尺度响应值后,对其进行 IDFT 得到对应的响应分数。如图 4 所示,尺度滤波器分数中间值最大,两端依次减小,最大响应分数对应的尺度就是最佳尺度。以平均区域重叠率(对每一帧预测目标矩形框与实际目标矩形框的重叠面积除以两个矩形框集合面积的值进行求和取平均)为衡量指标,插值后估计目标尺度比未进行插值估计尺度对应的重叠率提升 8% 左右。

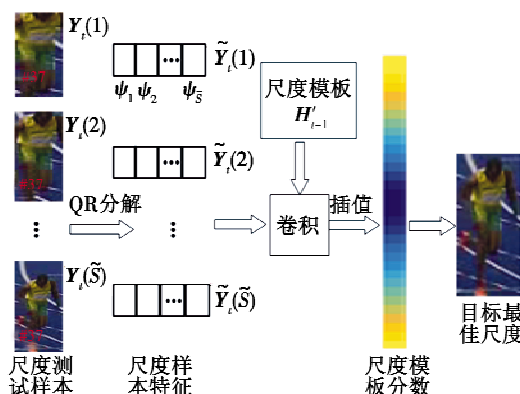


图 4 尺度滤波器分数

Fig. 4 Scale filter scores

### 2.3 改进的算法流程

本文算法采用 PCA 降维的方法提高位置滤波器的训练速度,通过压缩样本尺度信息来提高尺度滤波器的训练速度,目标的位置由位置滤波器分数和直方图分数的线性组合最大值来确定,目标的尺寸由插值后的尺度响应分数最大值来确定。算法流程如图 5 所示。

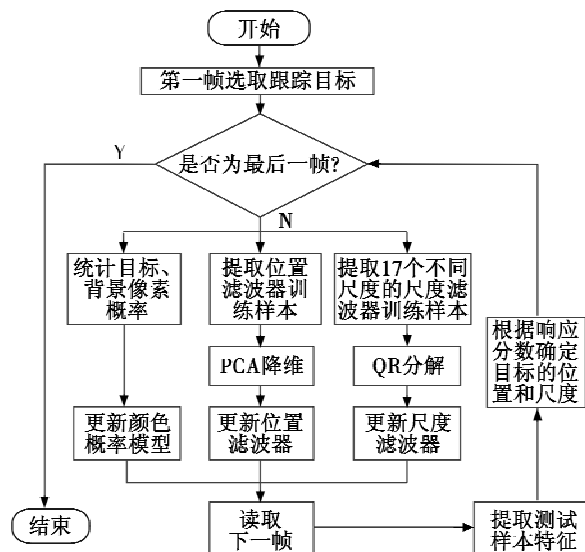


图 5 算法流程

Fig. 5 Flow chart of the algorithm

### 3 实验结果与分析

实验设备配置为 8 GiB RAM, Intel Corei5-7200U@2.5 GHz 处理器。本文算法和 Staple 算法进行对比实验所采用的视频来自 OTB2013 视频库,并在 Matlab R2016a 平台下编程实现。

#### 3.1 实验 1

实验 1 是在目标快速形变的环境下进行目标跟踪。采用公开的 Skiing 视频进行实验,从中选取几帧跟踪,结果如图 6 所示。

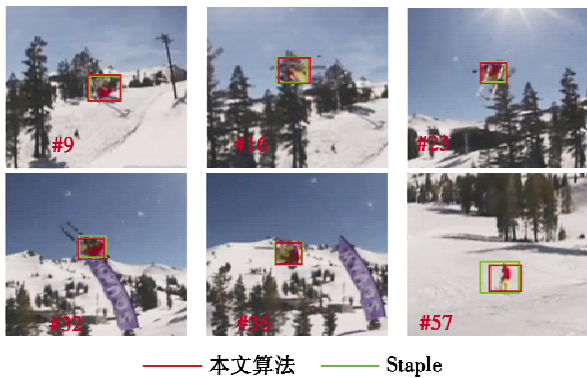


图 6 目标跟踪结果(Skiing)

Fig. 6 Results of target tracking(Skiing)

对 Skiing 视频每一帧中运动目标的实际位置坐标、Staple 算法预测目标位置坐标、本文算法预测目标位置坐标进行记录,结果如图 7 所示。

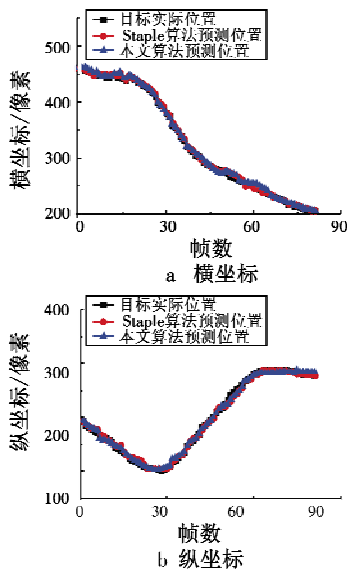


图 7 目标跟踪曲线(Skiing)

Fig. 7 Curves of target tracking(Skiing)

#### 3.2 实验 2

实验 2 是在背景目标颜色相似干扰、运动模糊的环境下进行目标跟踪。采用视频 CarDark 进行实验,从中选取几帧跟踪,结果如图 8 所示。



图 8 目标跟踪结果(CarDark)

Fig. 8 Results of target tracking(CarDark)

对 CarDark 视频每一帧中运动目标的实际位置坐标、Staple 算法预测目标位置坐标、本文算法预测目标位置坐标进行记录,结果如图 9 所示。

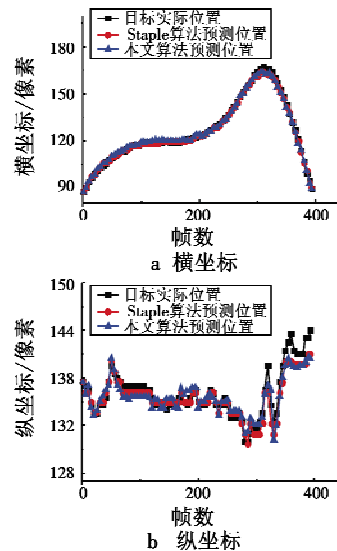


图 9 目标跟踪曲线(CarDark)

Fig. 9 Curves of target tracking(CarDark)

#### 3.3 实验结果分析

在每一个视频中进行目标跟踪时,对每一帧目标跟踪所需要的时间都进行了记录,如图 10 所示。

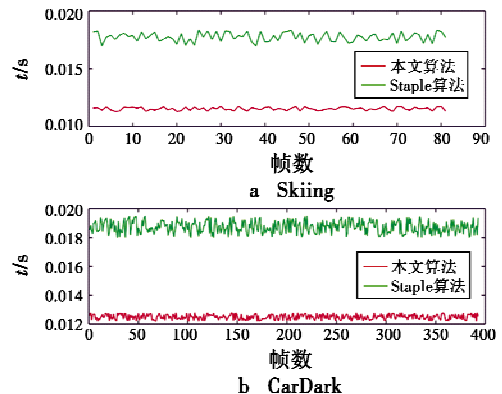


图 10 运行时间

Fig. 10 Running time of the two algorithms

为了定性评判跟踪结果的准确度,采用平均中心误差(对每一帧目标预测中心位置与目标实际中心位置之间的偏差进行求和取平均)和平均区域重叠率两个指标进行评估。从 OTB2013 中选取 10 个视频的实验结果进行对比,本文算法和 Staple 算法各自的指标结果如表 1 所示。

表 1 平均中心误差和区域重叠率

Table 1 The mean center error and region overlapping rate

视频	指标	本文算法	Staple 算法
Basketball	中心误差/像素	18.3	16.8
	区域重叠率/%	85.6	87.9
Bolt	中心误差/像素	4.3	4.6
	区域重叠率/%	100	100
CarDark	中心误差/像素	2.3	1.2
	区域重叠率/%	84.8	87.7
Dudek	中心误差/像素	11.5	13.6
	区域重叠率/%	79.8	79.1
Dog1	中心误差/像素	4.7	5.3
	区域重叠率/%	83.1	79.7
Jumping	中心误差/像素	8.4	6.2
	区域重叠率/%	66.2	70.6
Skiing	中心误差/像素	3.6	3.8
	区域重叠率/%	52.4	48.9
Tiger1	中心误差/像素	8.2	8.5
	区域重叠率/%	97.9	97.5
Tiger2	中心误差/像素	7.9	10.4
	区域重叠率/%	91.2	89
Trellis	中心误差/像素	3.6	3
	区域重叠率/%	97.4	99.3

对记录的时间进行统计计算,得到本文算法和 Staple 算法在实验中的平均帧频,结果如表 2 所示。

表 2 平均帧频

Table 2 Mean FPS 帧/s

视频	本文算法	Staple 算法	视频	本文算法	Staple 算法
Basketball	90.6	57.3	Jumping	111.4	70.9
Bolt	88.2	59.3	Skiing	86.2	55.6
CarDark	80.3	52.4	Tiger1	72.9	50.7
Dudek	73.8	49.6	Tiger2	74.4	50.1
Dog1	104.7	66.5	Trellis	83.3	55.9

由表 1、表 2 可以看出,本文算法和 Staple 算法在目标形变、背景和颜色相似干扰、运动模糊、目标尺度变化的场景下都能够准确地对目标进行跟踪,且两个算法的准确率相差很小。在运算速度上,本文算法比 Staple 算法快近 50%。

## 4 结束语

本文算法在 Staple 算法基础上,使用降维后的样本对位置滤波器和尺度滤波器进行训练,在跟踪准确率变化很小的情况下,速度有了明显提升。本文算法的不足在于有严重遮挡的情况下,不能准确地对目标进行跟踪,这也是接下来需要改进的地方。

## 参考文献

- [1] 蔡荣太,吴元昊,王明佳,等. 视频目标跟踪算法综述 [J]. 电视技术,2010,34(12):135-138,142.
- [2] DANELLJAN M, KHAN F S, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking [C]// Computer Vision and Pattern Recognition, IEEE, 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] ZHANG K H, ZHANG L, YANG M H, et al. Fast tracking via spatio-temporal context learning [J]. Computer Science, 2013(8):1311-1323.
- [5] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]//European Conference on Computer Vision, 2012:702-715.
- [6] 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, 2016:472-488.
- [7] 杨双祥,黄山. 提取目标不变矩的压缩感知跟踪 [J]. 电光与控制,2018,25(7):54-57,67.
- [8] DANELLJAN M, HÄGER G, KHAN F S, et al. Accurate scale estimation for robust visual tracking [C]//British Machine Vision Conference, 2014:651-662.
- [9] POSSEGGGER H, MAUTHNER T, BISCHOF H. In defense of color-based model-free tracking [C]//Computer Vision and Pattern Recognition, IEEE, 2015:2113-2120.
- [10] BERTINETTO L, VALMADRE J, GOLODETZ S, et al. Staple: complementary learners for real-time tracking [C]//Computer Vision and Pattern Recognition, IEEE, 2016:1401-1409.