

基于核相关的尺度自适应视觉跟踪

廖秀峰¹, 侯志强^{1,2}, 余旺盛¹, 王姣尧¹, 陈传华¹

¹空军工程大学信息与导航学院, 陕西 西安 710077;

²西安邮电大学计算机学院, 陕西 西安 710121

摘要 针对视觉跟踪中目标尺度变化对准确跟踪的不利影响, 提出一种基于核相关的尺度自适应视觉跟踪算法。首先, 通过建立核岭回归模型构建二维核相关定位滤波器, 采用融合后的多通道特征对滤波器进行训练, 提高目标定位的精度; 然后, 对目标区域进行多尺度采样, 样本缩放后提取其特征, 并构造为一维特征, 以此构建一维核相关尺度滤波器, 估计出目标的最佳尺度。在 OTB2013 平台上的实验结果表明, 与 8 种当前主流的跟踪算法相比, 本文算法的跟踪精度和成功率均有优势。在尺度变化条件下, 本文算法在快速准确跟踪的同时, 较好地实现了对目标尺度的自适应跟踪。

关键词 机器视觉; 尺度估计; 核岭回归模型; 特征融合

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/AOS201838.0715002

A Scale Adapted Tracking Algorithm Based on Kernelized Correlation

Liao Xiufeng¹, Hou Zhiqiang^{1,2}, Yu Wangsheng¹, Wang Jiaoyao¹, Chen Chuanhua¹

¹Information and Navigation Institute of Air Force Engineering University, Xi'an, Shaanxi 710077, China;

²School of Computer Science & Technology, Xi'an University of Posts & Telecommunications,
Xi'an, Shaanxi 710121, China

Abstract In order to solve the problem of accurate tracking and scale estimation in videos where targets change their scales, we propose a scale adapted tracking algorithm based on kernelized correlation. Firstly, we establish kernel ridge regression model and construct a two-dimensional kernelized correlation location filter. The center location of target is determined precisely by using fused multi-channel features. Then, the multi-scale samples of target area are obtained and their sizes are reset to the same with the model. By extracting their features and reconstructing to one-dimensional vector, we construct the one-dimensional kernelized scale filter to achieve optimal scale estimation. The experimental results on OTB2013 platform, especially on the scale changing benchmark dataset indicate that the proposed algorithm performs better in precision and success rate in comparison with eight mainstream tracking algorithms. Meanwhile, this algorithm can not only achieve an adapted tracking to the scale changing of target, but also locate its position fast and effectively.

Key words machine vision; scale estimation; kernel ridge regression model; feature fusion

OCIS codes 150.0155; 100.4999; 100.2960; 110.4155

1 引 言

在计算机视觉领域, 视觉跟踪一直都是研究热点和难点^[1-3], 在视频监控^[4]、成像制导^[5]、智能交通^[6]等领域都有着非常广泛的应用。近年来, 随着

相关研究的不断深入以及计算机软硬件性能的提升, 跟踪算法的性能得到了显著提高。但在实际跟踪过程中, 由于受到目标形变、尺度变化、光照变化以及遮挡等因素的影响, 精确快速的视觉跟踪系统的设计仍是一项具有挑战性的任务。

收稿日期: 2017-11-08; 收到修改稿日期: 2018-01-29

基金项目: 国家自然科学基金(61473309, 61703423, 41601436)、陕西省自然科学基金基础研究计划(2016JM6050)

作者简介: 廖秀峰(1993—), 男, 硕士研究生, 主要从事计算机视觉、机器学习等方面的研究。

E-mail: Liao.xf@foxmail.com

导师简介: 侯志强(1973—), 男, 博士, 教授, 博士生导师, 主要从事计算机视觉、模式识别等方面的研究。

E-mail: hou-zhq@sohu.com(通信联系人)

近年来,相关滤波方法取得了很大的突破,被广泛应用于视觉跟踪领域。Bolme 等^[7]首次将相关滤波方法应用到目标跟踪中,提出了 MOSSE 跟踪算法,算法在灰度图像上学习一个相关滤波器使得平方误差和最小,其跟踪速度可达 600 frame · s⁻¹。Henriques 等^[8]提出 CSK 跟踪算法,通过对样本进行循环移位获得大量样本,利用循环结构的性质和傅里叶变换在频域的快速计算使得训练出的滤波器更加稳健和精确。在此基础上,Danelljan 等^[9]提出了 CN 跟踪算法,使用颜色属性扩张 CSK 跟踪器,通过降维颜色属性特征实现了自适应颜色跟踪。Henriques 等^[10]对 CSK 算法作改进,使用方向梯度直方图(HOG)特征替代灰度特征,提出了具有多通道相关滤波器的 KCF 算法,提升了跟踪性能。

为解决相关滤波算法无法实现尺度自适应跟踪的问题,一些尺度估计方法^[11-15]逐渐被引入到跟踪算法中。Danelljan 等^[11]提出一种基于尺度金字塔滤波的尺度估计方法,较为准确地估计出目标尺度;王鑫等^[12]将尺度金字塔滤波方法与多层卷积特征结合,得到不错的跟踪及尺度估计效果;沈秋等^[13]将尺度金字塔滤波方法与多特征自适应选择相结合,同样实现了较好的尺度估计。为进一步提高目标定位和尺度估计的准确度,本文在 KCF 算法的基础上,提出一种基于核相关滤波器的尺度估计方法。该算法一方面对跟踪特征进行了改进,将方向梯度直方图特征^[16]与 HSI 模型的颜色信息进行融合得到新的特征,提高了目标定位的准确性;另一方面构造了一维核相关滤波器,利用尺度池提取多尺度特征并拉伸构造为一维特征向量,通过核相关估计出目标的最佳尺度,以实现目标尺度的自适应跟踪。为验证本文算法的有效性,利用 OTB2013^[17]评估基准的 51 组测试视频序列,进行整体性能评估,筛选出带有尺度变化属性的 28 组测试视频序列用于评估本文算法的尺度变化自适应性能,并与 KCF 等 8 种主流算法进行对比。

2 核岭回归跟踪模型

核岭回归(KRR)^[18]是一种以结构风险最小化为学习准则,解决在原始样本空间中不能用线性方法求解的非线性问题的算法,该算法具有较高的泛化能力。

对于现有的 m 个训练样本和标签 $(x_1, y_1), \dots, (x_m, y_m)$, 样本表示为一维向量 $x_i = (a_{i1}, a_{i2}, \dots, a_{in})^T$, y_i 为标签值,可以通过最小化其

正则化风险训练一个分类滤波器。一个线性的分类滤波器通常可以表示为

$$f(x) = \omega^T x + \xi, \quad (1)$$

式中 $\omega = (\omega_1, \omega_2, \dots, \omega_n)^T$ 是分类器参数, ξ 表示预测值与真实值的偏差,通常服从标准正态分布 $\xi \in N(0, \sigma^2)$ 。

为得到最优的分类器参数,首先构造损失函数 $L[y_i, f(x_i)]$,最优化问题可表示为

$$\min_{\omega, \xi} \left\{ \sum_{i=1}^m L[y_i, f(x_i)] + \lambda \|\omega\|_2^2 \right\}, \quad (2)$$

式中 $\|\omega\|_2^2 = \omega^T \omega$ 为正则化项, λ 是正则化参数,避免分类器过拟合。采用最小二乘方法(岭回归)构造损失函数,最优化解为

$$\omega = (X^T X + \lambda)^{-1} X^T Y, \quad (3)$$

式中 $X = (x_1, x_2, \dots, x_m)$ 是现有样本构成的矩阵, $Y = (y_1, y_2, \dots, y_m)^T$ 是样本的标签值构成的矩阵。

当在高维特征空间进行分类时,核函数可以将线性问题转换为核空间的非线性问题,从而提高分类效果。定义核相关 $k(x, x') = \langle \varphi(x), \varphi(x') \rangle$, $\langle \cdot, \cdot \rangle$ 表示两者的点积。

根据表示定理^[18], ω 可由 $\varphi(x)$ 的线性组合表示: $\omega = \sum_{i=1}^m \alpha_i \varphi(x_i)$, 此时求解 ω 的过程就变成了求解 α 的过程:

$$\alpha = (K^{xx} + \lambda)^{-1} Y, \quad (4)$$

式中 K^{xx} 为相关矩阵 $X^T X$ 的核化表达,其元素为 $K_{i,j}^{xx} = \langle \varphi(x_i), \varphi(x_j) \rangle$ 。

相关滤波器的训练需要大量样本作支撑,而大量的样本又将影响滤波器的计算速度,这对于样本的获取和滤波器的训练方式提出了要求,需要在获取足够数量的样本的同时采用某种训练方式来保证滤波器的精度和速度。循环矩阵的性质为这个问题提供了解决思路。将样本的循环作为训练样本,可以通过一次采样得到多个样本,解决了样本采样的问题;同时,循环结构的样本构成的样本循环矩阵可以利用循环矩阵的性质在傅里叶域中进行快速计算,从而破解样本数量和计算速度的矛盾。

对于一个样本及其标签 (x, y) , $x, y \in \mathbf{R}^{n \times 1}$, 如果 $x = (a_1, a_2, \dots, a_n)^T$ 的循环仍然是一个样本,那么可以构造样本矩阵 X , X 由样本 x 循环移位生成,因此是一个循环矩阵。

对于一个新的样本 $x' = (a'_1, a'_2, \dots, a'_n)^T$, x' 与 x 的所有循环样本的内积,根据卷积定理可以表示为

$$C(\mathbf{x})\mathbf{x}' = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ a_n & a_1 & \cdots & a_{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ a_n & a_{n-1} & \cdots & a_1 \end{bmatrix} \begin{bmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_n \end{bmatrix} = \mathcal{F}^{-1} [\mathcal{F}^*(\mathbf{x}) * \mathcal{F}(\mathbf{x}')], \quad (5)$$

式中 $C(\mathbf{x})$ 是由 \mathbf{x} 构成的循环矩阵, \odot 表示矩阵的 Hadamard 积; \mathcal{F}^{-1} 和 \mathcal{F} 分别表示逆离散傅里叶变换 (DFT) 以及 DFT, $*$ 表示共轭复数。

由于核相关矩阵 \mathbf{K}^{xx} 由样本核相关项 $k^{xx} = k(\mathbf{X}, \mathbf{x}')$ 生成, 是一个循环矩阵, 而循环矩阵的线性运算和逆矩阵也是循环矩阵, 因此矩阵 $(\mathbf{K}^{xx} + \lambda)^{-1}$ 也是循环矩阵。根据 (4) 式和 (5) 式, 核相关分类滤波器参数 α 可在频域快速计算:

$$\alpha = \mathcal{F}^{-1} \left[\left(\frac{1}{\hat{\mathbf{k}}^{xx} + \lambda} \right)^* \odot \hat{\mathbf{Y}} \right], \quad (6)$$

式中 \cdot^* 表示变量的傅里叶变换式, 核相关项 k^{xx} 为对称向量, 而对称向量的傅里叶变换为实数, 因此共轭符号可取消掉。

将候选样本 \mathbf{z} 构成的循环矩阵 \mathbf{Z} 输入训练好的核相关分类滤波器, 计算出所有循环样本响应:

$$f(\mathbf{Z}) = \mathcal{F}^{-1} \left[\left\langle \frac{\hat{\mathbf{Y}}}{\hat{\mathbf{k}}(\mathbf{X}, \mathbf{x}) + \lambda}, \hat{\mathbf{k}}(\mathbf{X}, \mathbf{z}) \right\rangle \right], \quad (7)$$

式中 $k(\mathbf{X}, \mathbf{z}) = k^{xz}$ 为核相关矩阵 \mathbf{K}^{xz} 的第一行, 公式中右侧的分数线表示元素间的点除运算。

选择高斯核作为二维跟踪滤波器的核函数, 则样本核相关表示为

$$k^{xx'} = \exp \left[- \frac{\|\mathbf{x}\|^2 + \|\mathbf{x}'\|^2 - 2 \sum_{d=1}^D \mathcal{F}^{-1}(\hat{\mathbf{x}}^* \odot \hat{\mathbf{x}'})}{2\sigma^2} \right], \quad (8)$$

式中 D 为样本的通道数。

3 本文跟踪算法

针对视觉跟踪任务中的尺度变化问题, 提出一种基于特征核相关的尺度自适应的跟踪算法。算法提取样本的 HOG 特征和颜色信息, 构建二维定位滤波器, 得到更为准确的目标跟踪响应图。在此基础上对跟踪目标进行多尺度采样^[11], 提取特征构建一维核相关尺度滤波器, 计算得到最佳尺度, 从而解决跟踪目标尺度变化的问题。算法增加颜色特征提高了跟踪的准确度, 进行尺度估计提高了跟踪的成功率, 具有快速、准确和尺度自适应的特点。

3.1 基于核相关和特征融合的目标定位

3.1.1 核相关定位滤波

在视觉跟踪中训练跟踪滤波器对大小为 $m \times n$ 的感兴趣区域 (ROI) 进行循环采样, 可以得到 $m \times n$ 个图像样本, 提取其特征作为训练跟踪滤波器的样本 \mathbf{x}_T , 样本矩阵 \mathbf{X}_T 是一个二维循环矩阵, 其中元素为 $\mathbf{x}_{ij} = \mathbf{E}_i \mathbf{x}_T \mathbf{E}_j$, \mathbf{E}_i 和 \mathbf{E}_j 为行置换和列置换矩阵。根据第 2 节关于核相关分类滤波器的介绍, 可以计算出二维定位滤波器的参数:

$$\alpha_T = \mathcal{F}^{-1} \mathcal{F} [(\mathbf{K}_T^{xx} + \lambda_T)^{-1} \mathbf{Y}_T] = \mathcal{F}^{-1} \left(\frac{\hat{\mathbf{Y}}_T}{\hat{\mathbf{k}}_T^{xx} + \lambda_T} \right), \quad (9)$$

式中 k_T^{xx} 为 \mathbf{K}_T^{xx} 第一个元素 $k(\mathbf{X}, \mathbf{x})$, λ_T 为其正则化参数, \mathbf{Y}_T 为样本标签。

3.1.2 基于特征融合的目标定位

在跟踪过程中, 在第 t 帧图像中根据第 $t-1$ 帧图像跟踪结果的目标中心位置 p_t 及目标大小 (w_t, h_t) , 考虑背景填充区域 p_{adding} 大小, 可以确定出 ROI 的中心位置 p_t 和大小 $(w_t, h_t) \times (1 + p_{\text{adding}})$, 然后从当前帧中提取出 ROI 图像块作为候选样本, 利用双线性插值方法进行缩放, 使得到的图像样本的大小与样本模板一致, 提取处理后的图像块的特征 $f_T \in \mathbf{R}^{M \times N \times D}$ 作为候选样本 \mathbf{z}_T , M 、 N 和 D 分别表示特征的三个维数, 对应特征样本的宽度、高度和通道数。

特征 $f_T \in \mathbf{R}^{M \times N \times D}$ 由候选区域的 31 维 HOG 特征与像素的 HSI 颜色模型信息构成, 由于亮度信息已经用于提取 HOG 特征, 因此只选取 HS 通道信息与 HOG 特征在每个单元上进行连接, 从而得到新的融合特征。新特征包含了梯度和颜色信息, 信息量比 HOG 特征更大, 在跟踪过程中, 对于一些颜色鲜明的目标可以有效提高跟踪精度。

将候选样本输入训练好的二维定位滤波器, 计算出二维响应图:

$$\hat{f}(\mathbf{Z}_T) = \left\langle \frac{\hat{\mathbf{Y}}_T}{\hat{\mathbf{k}}(\mathbf{X}_T^{xz}, \mathbf{x}_T) + \lambda_T}, \hat{\mathbf{k}}(\mathbf{X}_T^{xz}, \mathbf{Z}_T) \right\rangle, \quad (10)$$

样本标签采用二维高斯函数给定:

$$y_{T,ij} = \exp \left[- \frac{(i-i')^2 + (j-j')^2}{\sigma_T^2} \right], \quad \forall i, j = 0, 1, \dots, n-1, \quad (11)$$

式中 (i', j') 为目标坐标; σ_T 为高斯函数的带宽。

此时, 利用下式定位当前帧中跟踪目标的中心

位置 $p_t = (x_t, y_t)$:

$$(x_t, y_t) = \underset{m, n}{\operatorname{argmax}} f(\mathbf{z}_{T_{mn}}), \quad (12)$$

式中 (m, n) 为样本循环矩阵 \mathbf{Z}_T 的元素位置, 对应样本 $\mathbf{z}_{T_{mn}}$, $\mathbf{z}_{T_{mn}}$ 为样本 \mathbf{z}_T 中心位置移动到 (m, n) 处产生的循环样本, $f(\mathbf{Z}_T)$ 为 $\hat{f}(\mathbf{Z}_T)$ 的傅里叶反变换。

3.2 基于核相关的目标尺度估计

在定位目标的中心位置 p_t 后, 利用已知的目标尺度大小 $s_{t-1} = (w_{t-1}, h_{t-1})$, 对目标区域进行多尺度采样, 通过提取多尺度样本的融合特征构建一维

尺度滤波器对目标尺度进行估计。

3.2.1 一维特征提取

一维尺度滤波器以尺度因子 $a = 1.03$ 的等比变化构造尺度池, 尺度池 S 包含 31 个尺度:

$$s_i = \hat{a}^{(i-16)}, \quad i = 1, 2, \dots, 31, \quad (13)$$

式中 s_i 表示尺度池中第 i 个尺度。除了等比变化的方法, 还可以使用尺度因子的等差变化来构造尺度池, 实验发现, 等比变化可以覆盖更大范围的尺度而且划分更加精细, 能够提高尺度估计的准确性, 因此采用尺度因子的等比变化构造尺度池。

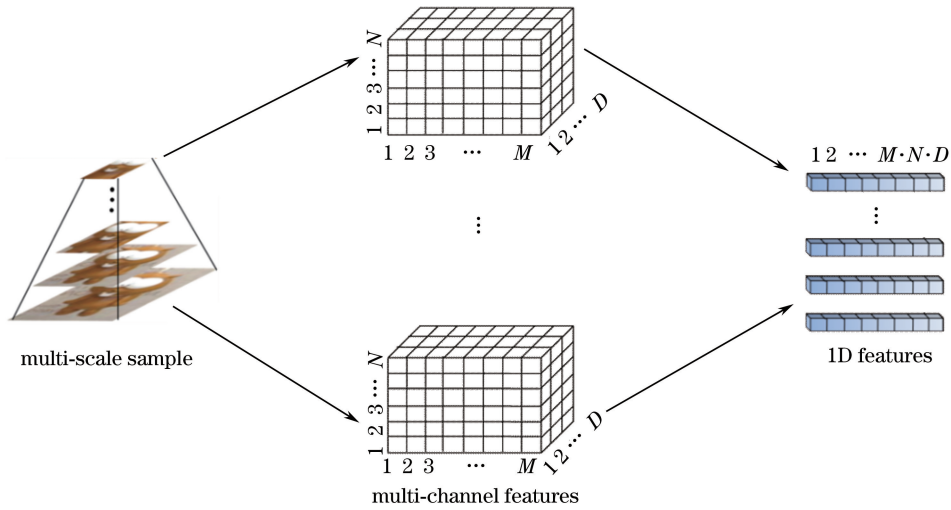


图 1 一维特征提取过程

Fig. 1 Extraction process of one-dimensional features

多尺度采样过程中, 对于第 i 个尺度等级, 以 p_t 为采样中心, 提取尺寸为 $s_i \cdot w_t \times s_i \cdot h_t$ 的图像块 J_i 。图 1 为一维特征的提取过程: 将每一个图像块 J_i 的 HOG 特征 $f_{s_i} \in \mathbf{R}^{M \times N \times D}$ 拉伸构造为一维特征, M, N, D 分别表示特征的长、宽和通道数, 然后把 31 个尺度样本的一维特征拼接为特征样本 $x_s \in \mathbf{R}^{M \times N \times D \times 31}$ 。

3.2.2 核相关尺度估计

采用一维高斯形式的样本标签

$$y_{s_i} = \exp\left[-\frac{(i-16)^2}{\sigma_s^2}\right], \quad \forall i = 1, 2, \dots, 31, \quad (14)$$

式中 σ_s 为一维尺度滤波器的标签函数带宽。由此得到一维尺度滤波器的分类器参数:

$$\boldsymbol{\alpha}_s = (\mathbf{k}_s^{xx} + \lambda_s)^{-1} \mathbf{Y}_s = \frac{\mathbf{Y}_s}{\mathbf{k}_s^{xx} + \lambda_s}, \quad (15)$$

式中 \mathbf{k}_s^{xx} 表示样本模板 \mathbf{x}_s 的自核相关, λ_s 为一维尺度滤波器的正则化参数。利用下式计算出尺度循

环样本的一维核相关响应:

$$f(\mathbf{Z}_s) = \mathcal{F}^{-1} \mathcal{F}[\langle \boldsymbol{\alpha}_s, \varphi(\mathbf{x}_s), \varphi(\mathbf{z}_s) \rangle] = \mathcal{F}^{-1} \langle \hat{\boldsymbol{\alpha}}_s^*, \hat{\mathbf{k}}_s^{xz} \rangle, \quad (16)$$

式中 \mathbf{Z}_s 为尺度循环样本, \mathbf{z}_s 为候选尺度样本。

寻找 $f(\mathbf{z}_{s_i})$ 中的最大响应值, 即可得到当前目标的最优尺度估计 s_t :

$$s_t = \hat{a}^{\lceil \operatorname{argmax}_i f(\mathbf{z}_{s_i})^{-16} \rceil}, \quad (17)$$

式中尺度等级 $i \in \left\{ \left\lfloor -\frac{S-1}{2} \right\rfloor, \dots, \left\lfloor \frac{S-1}{2} \right\rfloor \right\}$, \mathbf{z}_{s_i} 表示候选尺度样本中心位置循环至 i 处时产生的样本, 根据估计出的最佳尺度与上一帧目标大小计算当前的目标大小:

$$(w_t, h_t) = (w_{t-1}, h_{t-1}) \cdot s_t. \quad (18)$$

3.3 模型更新策略

在跟踪过程中, 受到目标的非刚性形变、尺度变化、光照变化等因素影响, 容易出现跟踪漂移, 因此跟踪过程中必须对跟踪模型进行更新。本文算法中, 模型的更新主要体现在核相关滤波器的参数模

板与样本模板的更新。

候选样本的定位滤波响应值反映候选样本与样本模板的匹配程度,因此,合理利用响应值调整模型更新策略,通过在低响应值的情况下加强模板中的首帧信息可以提高定位滤波器对物体遮挡的稳健性。本文算法提取最大响应值进行判定,对于二维定位滤波器,设定响应阈值 $\xi=0.2$,当 $\max(r_{es})>\xi$ 时,更新策略为

$$\begin{cases} \alpha_T^t = (1 - \eta_T)\alpha_T^{t-1} + \eta_T\alpha_{TN}^t \\ \mathbf{X}_T^t = (1 - \eta_T)\mathbf{X}_T^{t-1} + \eta_T\mathbf{X}_{TN}^t \end{cases}, \quad (19)$$

当 $\max(r_{es})<\xi$ 时,更新策略为

$$\begin{cases} \alpha_T^t = (1 - \eta_T)\alpha_T^{t-1} + \eta_T [(1 - \beta)\alpha_{TN}^t + \beta\alpha_{TN}^1] \\ \mathbf{X}_T^t = (1 - \eta_T)\mathbf{X}_T^{t-1} + \eta_T [(1 - \beta)\mathbf{X}_{TN}^t + \beta\mathbf{X}_{TN}^1] \end{cases}, \quad (20)$$

式中 r_{es} 表示候选样本的滤波响应, α_T^{t-1} 和 \mathbf{X}_T^{t-1} 分别为其在第 $t-1$ 帧时的滤波器参数和跟踪模板,

表 1 基于核相关的尺度自适应视觉跟踪算法

Table 1 Scale adapted tracking algorithm based on kernelized correlation

Input:	Image sequence: I_1, I_2, \dots, I_n . Initial target position: $p_0 = (x_0, y_0)$, and initial target scale: $s_0 = (w_0, h_0)$
Output:	The estimated position of target: $p_t = (x_t, y_t)$, and estimated scale: $s_t = (w_t, h_t)$
for	$t=1, 2, 3, \dots, n$, do:
1	Locate the ROI area in frame # t centered at p_{t-1} with the scale of s_{t-1} ;
2	Crop out the ROI image and resize to the size of sample template;
3	Extract the HOG and color features;
4	Learn the kernelized correlation response map using Eq.(10);
5	Locate the center of the target p_t in frame # t using Eq.(12);
6	Obtain the multi-scale sample image $I_s = \{I_{s_1}, I_{s_2}, \dots, I_{s_S}\}$ in frame # t based on p_t and s_{t-1} ;
7	Build scale filters by extracting fusion features from the above multi-scale image;
8	Compute the kernelized correlation response score using Eq.(16);
9	Estimate the optimal scale s_t of target in the frame # t using Eq.(17) and Eq.(18);
10	Update the translation filters using Eq.(19) and Eq.(20);
11	Update the scale filters using Eq.(21).
Until	End of the image sequence.

4 实 验

为充分验证本文算法的有效性,采用 MATLAB 2014a 进行编程实现。将本文算法在 Intel Xeon 2.4 GHz 处理器上进行测试,并采用 GPU(TITANX)进行加速,算法参数设置如下:目标 ROI 区域大小为目标的 2.5 倍,即 $p_{\text{adding}}=1.5$;二维核相关跟踪滤波器参数与 KCF 跟踪算法保持一致,正则化参数 $\lambda_T=10^{-4}$,标签函数空间带宽 $\sigma_T=0.1$,高斯核带宽 $\sigma_p=0.5$,学习率 $\eta_T=0.02$;一维核相关尺度滤波器参数中,标签函数空间带宽 $\sigma_s=0.25$,正则化参数 $\lambda_s=10^{-2}$,学习率 $\eta_s=0.025$ 与 DSST 跟踪算法保持一致,而尺度池样本大小 $S=$

α_{TN}^t 和 \mathbf{X}_{TN}^t 分别表示在第 t 帧中得到的滤波器参数和样本, α_{TN}^1 和 \mathbf{X}_{TN}^1 分别表示首帧中的滤波器参数模板和样本模板, η_T 为定位滤波器学习率,表示在每一帧模板更新中新获得的样本模板和滤波器参数模板的更新权重, β 为首帧信息的更新权重。

对于一维尺度滤波器,本文算法不设定阈值判定, α_s^{t-1} 和 \mathbf{X}_s^{t-1} 分别表示第 $t-1$ 帧中尺度滤波器的参数模板和样本模板,则在第 t 帧中,尺度滤波器的更新策略为

$$\begin{cases} \alpha_s^t = (1 - \eta_s)\alpha_s^{t-1} + \eta_s\alpha_{SN}^t \\ \mathbf{X}_s^t = (1 - \eta_s)\mathbf{X}_s^{t-1} + \eta_s\mathbf{X}_{SN}^t \end{cases}, \quad (21)$$

式中 α_{SN}^t 和 \mathbf{X}_{SN}^t 表示第 t 帧中的尺度滤波器参数和样本, η_s 为尺度滤波器学习率。

3.4 算法流程

综上所述,本文算法主要跟踪流程归纳如表 1,算法整体流程图如图 2 所示。

31,尺度因子 $a=1.03$,高斯核带宽 $\sigma_s=1$,首帧信息权值 $\beta=0.6$,这些参数均是经过大量实验得出的效果较好的经验值。在 51 组视频实验中所有参数的设置保持不变。

为了充分说明本文算法在处理目标尺度变化与遮挡等方面的优势,算法在 OTB2013 数据平台上进行了 51 组视频实验,并与 8 种主流的跟踪算法进行了比较,展示了算法良好的整体性能。此外,为充分说明算法出色的尺度自适应能力,选取其中具有尺度变化的视频另外进行了精度和成功率的比较。选取的对比算法有 CNT^[19]、DLT^[20]、KCF^[10]、DSST^[11]、CN^[9]、CST^[21]、SST^[22]、NRMLC^[23],其中 CNT、DLT 是基于深度学习的跟踪算法,KCF、

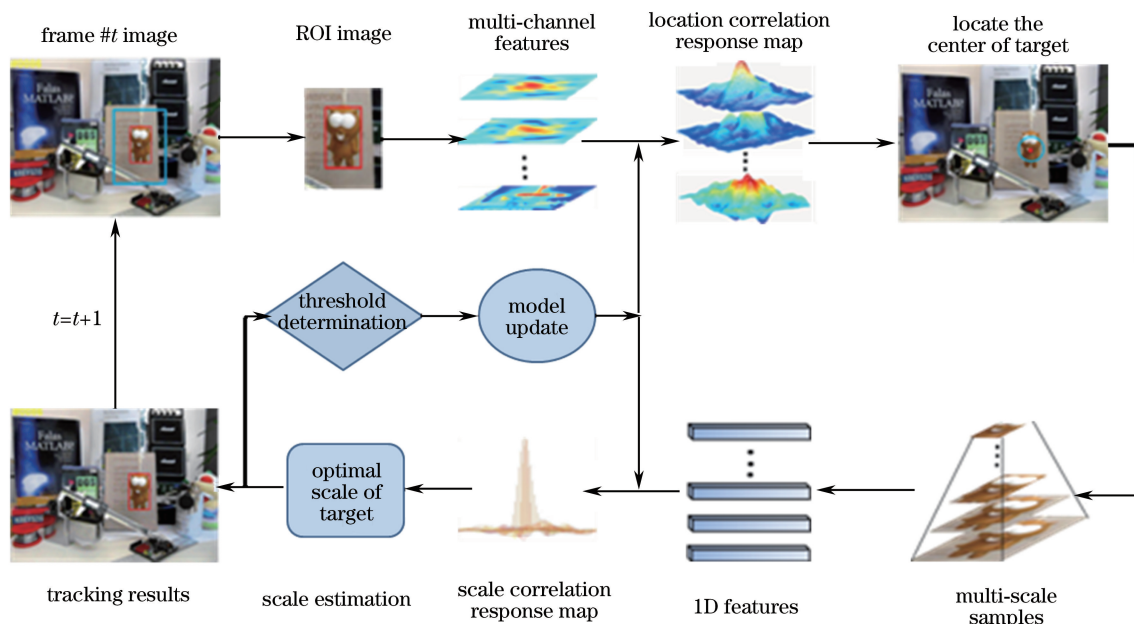


图 2 本文算法流程图

Fig. 2 Flow chart of proposed tracking method

DSST 和 CN 是基于相关滤波的跟踪算法, CST、SST 和 NRMLC 是近 3 年的主流跟踪算法, 跟踪结果由其作者提供。

4.1 定性分析

1) 尺度变换。尺度变化是目标跟踪中比较常见的问题, 由于目标尺度的变化导致固定大小的跟踪框无法获取完整的目标信息, 从而影响跟踪效果甚至丢失目标。以“car scale”“fleet face”“dog1”“singer2”为例, 这 4 组视频中的目标均出现一定程度的尺度变化, CST、DLT、CNT、SST、NRMLC、DSST 等算法都具有尺度适应能力并可以始终跟踪上目标。对于尺度变化的目标, 其本身颜色信息对目标尺度的估计有重要作用, 本文算法由于采用核相关滤波对尺度进行估计并且加入了颜色信息, 因而能够更好地适应目标的尺度变化。

2) 光照变化。光照变化常常导致目标的颜色信息发生明显变化, 因而容易出现跟踪丢失或偏差。以“skating1”和“shaking”为例, 跟踪过程中背景光照发生剧烈变化, 要求算法对光照具有较强的稳健性。比如在“shaking”中, 由于所采用的特征对光照变化的适应能力不强, 随着目标运动, CST、KCF、CNT、NRMLC 算法跟踪漂移直至丢失目标, 而 HOG 特征对光照具有良好的稳健性, 本文算法结合了 HOG 特征和颜色信息, 可以较好地跟踪目标。

3) 目标形变。目标外观发生变化会导致当前

帧信息与模板匹配程度降低, 增加跟踪难度。以“sylvestor”和“basketball”为例, 目标外观都发生明显变化, 比如“sylvestor”中目标在三维空间旋转, “basketball”中目标跑动间形体变化, CST、SST 算法均跟踪失败, 而本文算法与 ACT 算法跟踪特征中包含了目标的颜色信息, 跟踪效果相对较好。在模板更新方面, 本文算法对响应值设定了阈值判定, 增加了首帧信息, 使得跟踪更加稳定。

4) 目标快速运动。快速运动常常导致图像模糊, 使得跟踪算法难以获得目标的表现信息而导致跟踪丢失。以“tiger1”和“freeman1”为例, 目标快速移动导致 NRMLC、CNT、DLT、KCF、ACT 失效, 而本文算法和 CST、SST 算法由于对目标尺度估计准确, 获取到的特征更加准确, 对于运动模糊的快速运动目标具有较好的稳健性, 如“tiger1”第 56 帧。

5) 目标遮挡。目标遮挡使得跟踪算法不能获得足够的目标信息而导致跟踪偏差或丢失。以“coke”和“jogging-1”为例, 目标在跟踪过程中被不同程度地遮挡, 在“coke”第 277 帧中和“jogging-1”第 85 帧中, 目标被遮挡后再次出现, DLT、CNT、SST、DSST 算法出现了跟踪漂移, 而本文算法的更新策略对目标遮挡进行了判断, 在响应较低的情况下加入首帧信息, 避免了遮挡情况下坏样本对模板的破坏, 因此对目标遮挡问题具有较好的稳健性, 能够始终准确跟踪目标。

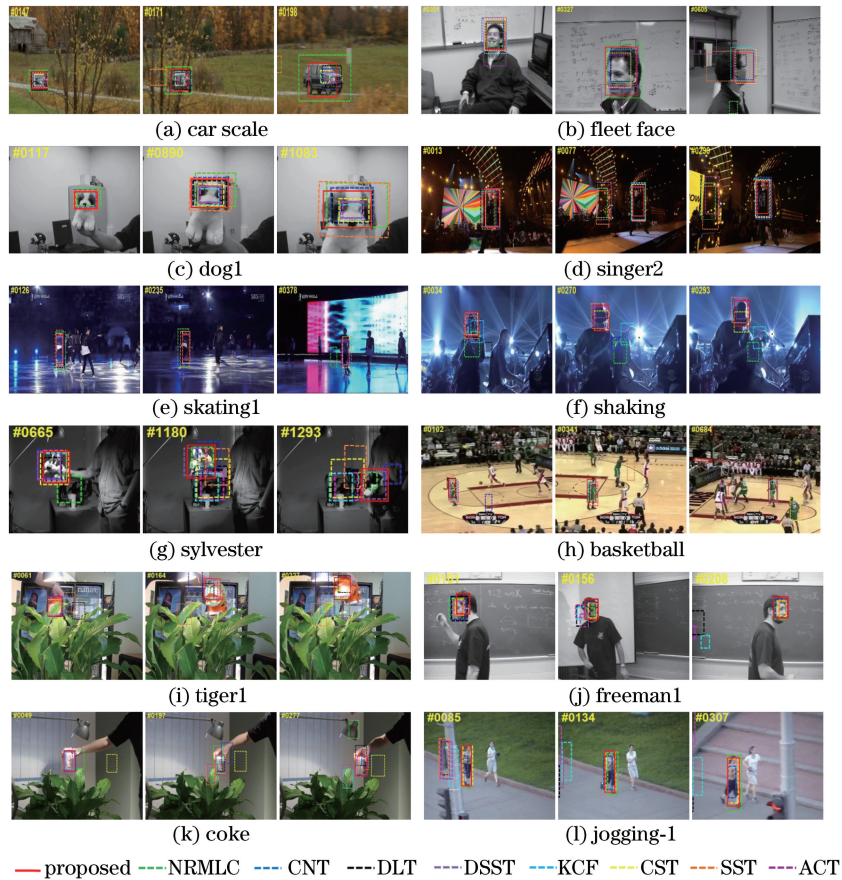


图 3 9 种算法的部分跟踪结果对比

Fig. 3 Qualitative comparison of tracking results of 9 trackers

4.2 定量分析

从算法对单个视频序列以及 51 组视频序列整体的跟踪性能两个方面对算法进行定量分析。

4.2.1 对单个视频的定量分析

针对所选取的 12 组视频序列,采用中心误差和覆盖率两个评价指标对算法进行对比分析。

平均中心误差是跟踪结果与真实目标中心位置之间的平均欧氏距离,体现算法跟踪的准确程度,平均中心误差越小,表明算法跟踪的精度越高。覆盖率是指跟踪结果和真实目标的重叠率,其值表征算法的尺度适应能力,值越大,说明尺度适应性越好,其表达式为

$$V_{\text{overlap}} = |S_T \cap S_{GT}| / |S_T \cup S_{GT}|, \quad (21)$$

式中 S_T 表示算法跟踪结果的目标区域, S_{GT} 是根据数据集中给定的目标真实中心位置与目标大小得到的真实目标区域,与算法跟踪得到的目标区域形成比较, $|\cdot|$ 表示计算区域面积,覆盖率定义为跟踪结果和真实目标交集面积大小与两者并集面积大小的比值。表 2 为本文算法与 8 种对比算法在 12 组测试视频上的中心误差和覆盖率,表中,括号内数值

为覆盖率的百分数,括号外数值为中心位置误差。从表 2 可以看出,相对其他算法,本文算法在不同视频中的覆盖率均能达到较高水平,表明本文算法对跟踪目标的尺度变化具有更好的适应能力,中心位置误差始终保持在较低水平,说明本文算法在跟踪精度方面较其他算法更有优势,也更能适应不同类型的视频。

4.2.2 算法综合性能的定量分析

为综合评判本文算法对尺度变化的适应能力以及对所有测试视频的跟踪性能,分别对 OTB2013 中 28 组具有尺度变化属性的测试视频以及 OTB2013 的 51 组测试视频进行测试分析。

每一帧视频中,当满足 $V_{\text{overlap}} > t_0$ 时,则认为目标跟踪成功,其中 t_0 为给定的成功率阈值。以跟踪精度和跟踪成功率作为评价指标进行定量分析,跟踪成功率为算法在整个视频中成功跟踪的帧数与视频总帧数的比值,跟踪精度表示算法在整个视频中跟踪结果的平均中心位置误差小于给定的阈值时,算法成功跟踪的帧数与视频总帧数的比值。

图 4 是本文算法与 8 种对比算法在 28 组具有

表 2 12 组测试序列的中心位置误差和覆盖率

Table 2 Center location errors and overlap rates of 12 sequences

Sequence	Proposed	CNT	DSST	CST	SST	KCF	NRMLC	DLT	CN
car scale	7.3(75)	23.4(57)	18.8(46)	12.5(48)	87(59)	16.1(47)	10.4(74)	25.5(61)	25.2(43)
fleet face	22(69)	24.7(60)	28.3(58)	67.2(54)	60.8(67)	26.4(63)	96.9(44)	27.5(56)	126.2(26)
dog1	3.8(99)	6.8(95)	4.6(66)	4.8(67)	4.9(89)	4.1(64)	11.6(91)	4.4(92)	3.5(68)
singer2	7.2(100)	6.8(100)	8.2(100)	7.4(100)	175.2(4)	10.2(100)	195.4(3)	173.0(3)	167.1(4)
skating1	6.2(53)	6.7(60)	6.8(52)	7.9(58)	8.8(62)	7.7(50)	15(52)	52.9(49)	8.0(52)
shaking	7.7(74)	74.1(5)	8(73)	5.7(72)	8.1(75)	113.2(4)	109.4(3)	—	15.1(60)
sylvester	8.7(75)	10.7(62)	14.8(63)	12.9(68)	11.2(63)	13.3(67)	24.5(49)	10.9(51)	9.5(68)
basketball	5.3(78)	534.1(6)	111.6(28)	23.5(57)	106.0(22)	8.1(67)	60.0(7)	12.0(51)	9.3(63)
tiger1	12(68)	94.2(15)	19.5(63)	11.2(74)	93.5(16)	15.7(68)	54.5(22)	23.2(58)	61.2(20)
freeman1	7.4(59)	7.9(53)	112.5(24)	9.7(41)	9.8(37)	94.6(23)	7.4(49)	103.6(28)	159.9(22)
coke	10.5(65)	36.7(35)	12.7(60)	148.7(4)	25.9(45)	18.7(55)	62.1(17)	20.1(53)	30.8(42)
jogging-1	4.3(80)	6.2(62)	112(19)	3.9(81)	144.6(20)	87.9(19)	7.2(75)	113(18)	101.7(19)
Average	8.5(75)	69.4(51)	38.2(55)	26.3(61)	61.3(47)	34.7(52)	54.5(40)	51.5(45)	59.8(41)

尺度变化属性的测试视频上进行测试的精度曲线和成功率曲线,图例中的数值分别为平均中心位置误差为 20 pixel 时的精度数值和成功率曲线下的面积 (AUC) 值。从图中曲线和图例数值可以看出,针对 28 组具有尺度变化的测试视频,本文算法的精度和

成功率均领先于其他算法,精度相比于 KCF 算法提高了 9%,尤其是成功率比 KCF 算法提高了 25.1%。由于算法对尺度变化的适应性主要体现在成功率上,前述结果说明本文算法具有很好的尺度自适应能力。

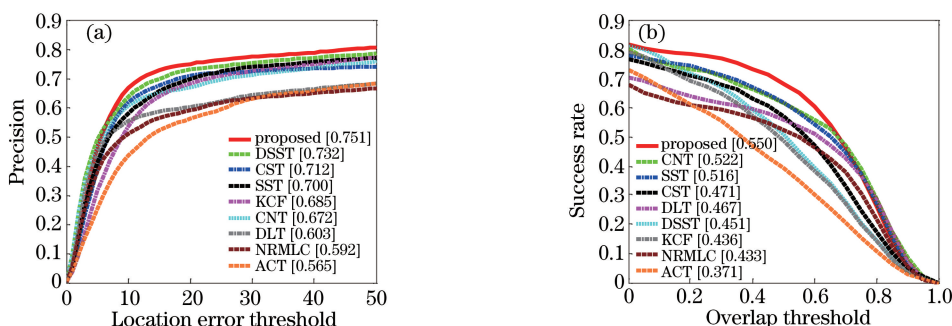


图 4 28 个尺度变化测试序列的 (a) 精度曲线和 (b) 成功率曲线

Fig. 4 (a) Precision plots and (b) success plots of 28 sequences with scale variations

为了进一步分析算法在不同条件下的跟踪性能,在 OTB2013 的 51 组测试视频上进行性能测试。图 5 为 9 种算法在 51 组测试视频上的总的精度曲线和成

功率曲线,表 3 和表 4 则分别列出了 11 种不同属性的跟踪条件下算法的跟踪精度和成功率。表格第一行是不同属性的缩写,括号内为该种属性包含的视频

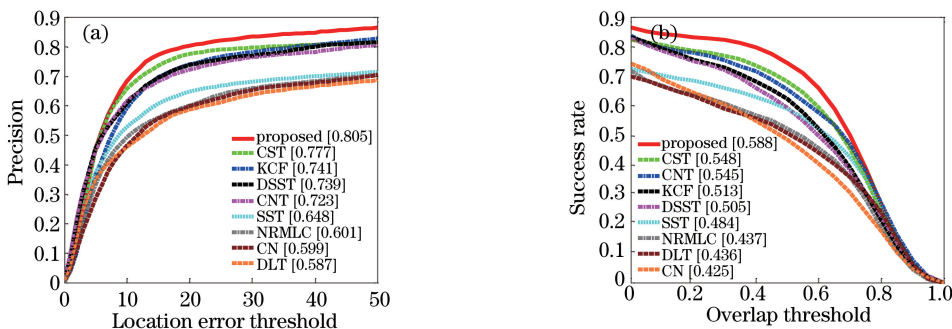


图 5 51 个测试序列的 (a) 精度曲线和 (b) 成功率曲线

Fig. 5 (a) Precision plots and (b) success plots of 51 sequences

个数,第一列为测试的算法,表格中间部分为不同算法在不同属性的测试视频下的精度值和成功率。11种属性分别为:尺度变化(SV),光照变化(IV),遮挡

(OCC),背景变化(BC),目标形变(DEF),运动模糊(MB),快速运动(FM),平面内旋转(IPR),离面转动(OPR),目标超出视野(OV),低分辨率(LR)。

表 3 不同属性下算法跟踪精度

Table 3 Tracking precision values on 11 different attributes

Algorithm	SV (28)	IV (25)	OCC (29)	BC (21)	DEF (19)	MB (12)	FM (17)	IPR (31)	OPR (39)	OV (6)	LR (4)
Proposed	0.744	0.757	0.792	0.756	0.831	0.624	0.630	0.779	0.791	0.648	0.516
CNT	0.662	0.566	0.662	0.646	0.687	0.507	0.500	0.661	0.672	0.502	0.557
DSST	0.740	0.741	0.725	0.691	0.657	0.603	0.562	0.780	0.732	0.533	0.534
CST	0.707	0.676	0.726	0.773	0.756	0.591	0.516	0.708	0.742	0.596	0.454
SST	0.688	0.603	0.588	0.644	0.487	0.408	0.425	0.630	0.599	0.406	0.527
KCF	0.680	0.729	0.749	0.752	0.741	0.650	0.602	0.725	0.730	0.649	0.379
NRMLC	0.597	0.437	0.583	0.497	0.541	0.378	0.397	0.511	0.546	0.492	0.542
DLT	0.590	0.534	0.574	0.495	0.563	0.453	0.446	0.548	0.561	0.444	0.396
CN	0.554	0.532	0.582	0.642	0.523	0.396	0.416	0.615	0.605	0.434	0.405

表 4 不同属性下算法跟踪成功率

Table 4 Tracking success rates on 11 attributes

Algorithm	SV (28)	IV (25)	OCC (29)	BC (21)	DEF (19)	MB (12)	FM (17)	IPR (31)	OPR (39)	OV (6)	LR (4)
Proposed	0.541	0.560	0.579	0.551	0.609	0.493	0.491	0.566	0.572	0.537	0.382
CNT	0.508	0.456	0.503	0.488	0.524	0.417	0.404	0.495	0.501	0.439	0.437
DSST	0.451	0.506	0.480	0.492	0.474	0.458	0.433	0.532	0.491	0.490	0.352
CST	0.466	0.486	0.506	0.567	0.551	0.474	0.411	0.496	0.514	0.509	0.349
SST	0.504	0.459	0.436	0.489	0.391	0.313	0.340	0.451	0.437	0.347	0.407
KCF	0.427	0.494	0.513	0.533	0.533	0.499	0.461	0.497	0.496	0.550	0.310
NRMLC	0.427	0.341	0.437	0.370	0.392	0.303	0.334	0.367	0.389	0.410	0.428
DLT	0.455	0.405	0.423	0.339	0.394	0.363	0.360	0.411	0.412	0.367	0.346
CN	0.363	0.390	0.404	0.453	0.388	0.329	0.334	0.437	0.418	0.410	0.311

由表 3 和表 4 可知,除了低分辨率的视频,本文算法对各种属性视频的跟踪精度和成功率均位于最优或次优。另外,从图 5 可以看出,本文算法对 51 组包含多种属性的测试视频整体跟踪精度和成功率优于其他算法。由此可见,对于各种属性的部分测试视频,以及 51 组测试视频整体而言,本文算法都

具有很好的跟踪性能。

4.3 算法跟踪速率

在 GPU 条件下,本文算法在 OTB2013 的 51 组实验视频中的跟踪速率为 17.3~136.1 frame·s⁻¹,平均跟踪速率为 56.3 frame·s⁻¹,在 CPU 条件下速度则相对慢一些,平均跟踪速率为 43.2 frame·s⁻¹。

表 5 本文算法与 8 种算法跟踪速率对比

Table 5 Tracking speed comparison of 9 trackers

Tracker	Proposed	CNT	DSST	CST	SST	KCF	NRMLC	DLT	CN	
Code	M	M	M	M	M+C	M	M	M	M	
Platform	CPU	CPU+GPU	CPU	CPU	CPU	CPU+GPU	CPU	CPU	CPU+GPU	CPU
Tracking speed / (frame·s ⁻¹)	43.2	56.3	5	24	2.2	2.2	172	1.28	15	—

Note: M, MATLAB; C, C++

表 5 列出了本文算法与 8 种当前的主流跟踪算法的跟踪速率对比,分别列出了各个算法的编程方式和实验平台,“—”表示文中没有给出算法的平均速

率。从表 5 可以看出,与传统算法和深度学习等比较新的算法相比,本文算法在跟踪速度上占有较大优势,但由于加入了尺度估计,速度要慢于 KCF 算法。

5 结 论

在 KCF 算法的基础上提出一种基于多维特征核相关的尺度自适应跟踪算法。该算法利用 KCF 算法的跟踪滤波器进行跟踪,并改进了跟踪滤波器使用的特征,将 HOG 特征与颜色信息相结合,提高跟踪精度。同时,为解决目标跟踪过程中尺度变化的问题,基于核相关理论,提取目标多尺度的 HOG 特征构建尺度滤波器,进行目标的精确尺度估计。实验结果表明,在复杂跟踪条件下,本文算法不仅能够准确跟踪目标,同时较好地解决了跟踪过程中目标尺度变化的问题,显著提高跟踪的成功率。

在实验中还发现,当目标出现长时间大面积遮挡后再次出现时,本文算法容易出现跟踪漂移,导致无法重新捕获目标。为解决这一问题,考虑将检测思想加入算法中,提高遮挡条件下的跟踪性能,这将是下一步研究的重点。另外,为进一步提高跟踪精度,采用深度学习提取并改进特征也将是下一步研究工作的另一个重点方向。

参 考 文 献

- [1] Smeulders A, Chu D, Cucchiara R, *et al.* Visual tracking: an experimental survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7): 1442-1468.
- [2] Hou Z Q, Han C Z. A survey of visual tracking[J]. Acta Automatica Sinica, 2006, 32(4): 603-617.
侯志强, 韩崇昭. 视觉跟踪技术综述[J]. 自动化学报, 2006, 32(4): 603-617.
- [3] Wang N Y, Shi J P, Yeung D, *et al.* Understanding and diagnosing visual tracking systems [C]. IEEE International Conference on Computer Vision, 2015: 3101-3109.
- [4] Migliore D, Matteucci M, Naccari M. A reevaluation of frame difference in fast and robust motion detection [C]. International Workshop on Video Surveillance and Sensor Networks, 2006: 215-218.
- [5] Basafa E, Foroughi P, Hossbach, *et al.* Visual tracking for multi-modality computer-assisted image guidance [C]. SPIE Medical Imaging, 2017: 101352S.
- [6] Bardet F, Chateau T. MCMC particle filter for real-time visual tracking of vehicles[C]. IEEE Conference on Intelligent Transportation Systems, 2008: 539-544.
- [7] Bolme D, Beveridge J, Draper B, *et al.* Visual object tracking using adaptive correlation filters[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2010: 2544-2550.
- [8] Henriques J, Caserio R, Martins P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels [C]. European Conference on Computer Vision and Pattern Recognition, 2010: 702-715.
- [9] Danelljan M, Khan F, Felsberg M, *et al.* Adaptive color attributes for real-time visual tracking [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2014: 1090-1097.
- [10] Henriques J, Caserio 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.
- [11] Danelljan M, Hager G, Khan F, *et al.* Accurate scale estimation for robust visual tracking[C]. British Machine Vision Conference, 2014.
- [12] Wang X, Hou Z Q, Yu W S, *et al.* A scale adaptive robust tracking algorithm based on fusion of multilayer convolutional features [J]. Acta Optica Sinica, 2017, 37(11): 1115005.
王鑫, 侯志强, 余旺盛, 等. 基于多层卷积特征融合的目标尺度自适应鲁棒跟踪[J]. 光学学报, 2017, 37(11): 1115005.
- [13] Shen Q, Yan X L, Liu L F, *et al.* Multi-scale correlation filter tracker based on adaptive feature selection [J]. Acta Optica Sinica, 2017, 37(5): 0515001.
沈秋, 严小乐, 刘霖枫, 等. 基于自适应特征选择的多尺度相关滤波跟踪[J]. 光学学报, 2017, 37(5): 0515001.
- [14] Li S S, Zhao G P, Wang J Y, *et al.* Distractor-aware object tracking based on multi-feature fusion and scale-adaption[J]. Acta Optica Sinica, 2017, 37(5): 0515005.
李双双, 赵高鹏, 王建宇, 等. 基于特征融合和尺度自适应的干扰感知目标跟踪[J]. 光学学报, 2017, 37(5): 0515005.
- [15] Yu W S, Tian X H, Hou Z Q, *et al.* Multi-scale mean shift tracking[J]. IET Computer Vision, 2015, 9(1): 110-123.
- [16] Navneet D, Bill T. Histograms of oriented gradients for human detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2005: 886-893.
- [17] Wu Y, Lim J, Yang M H. Online object tracking [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2013: 2411-2418.
- [18] Rifkin R, Yeo G, Poggio T. Regularized least-squares classification [J]. Acta Electronica Sinica, 2007, 190(1): 93-104.
- [19] Zhang K H, Liu Q S, Wu Y. Robust visual tracking

- via convolutional networks without training[J]. IEEE Transactions on Image Processing, 2015, 25 (6): 1779-1792.
- [20] Wang N Y, Yeung D. Learning a deep compact image representation for visual tracking [C]. Advances in Neural Information Processing Systems, 2013: 809-817.
- [21] Zhang T Z, Bibi A, Ghanem B. In defense of sparse tracking: circulant sparse tracker [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 3880-3888.
- [22] Zhang T, Liu S, Xu C S, *et al.* Structural sparse tracking[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 150-158.
- [23] Liu F, Zhou T, Yang J, *et al.* Visual tracking via nonnegative regularization multiple locality coding [C]. IEEE International Conference on Computer Vision, 2016: 72-80.