

文章编号:1674-2869(2019)01-0098-05

# 优化搜索策略的KCF目标跟踪算法

杨志方, 陈曦

武汉工程大学电气信息学院, 湖北 武汉 430205

**摘要:**针对核相关滤波跟踪算法存在实时性较差的问题,提出了一种优化搜索策略的改进算法。首先,在检测到随机选取的视频某一帧中目标中心位置后,计算该目标图像块的均值和标准差。再设定一个排序队列以及两个自适应阈值来筛选一些特征与目标差异较大的候选块。在视频下一帧中,均值与标准差的差值小于设定阈值的候选块会优先检测并计算响应。实验结果表明,改进后的算法与原算法相比帧率提升可达10%左右,且跟踪精度较KCF、CSK、Struc等其它算法提升2.2%、14.4%和24.9%。

**关键词:**目标跟踪;相关滤波;候选图像块;搜索策略

中图分类号:TP391 文献标识码:A doi:10.3969/j.issn.1674-2869.2019.01.017

## Optimized Searching Strategy for KCF Object Tracking Algorithm

YANG Zhifang, CHEN Xi

School of Electrical and Information Engineering, Wuhan Institute of Technology, Wuhan 430205, China

**Abstract:** To improve the real-time performance of kernelized correlation filters (KCF) algorithm, we proposed an optimized searching strategy for KCF tracking algorithm. The center position of target in a randomly selected video frame was detected, and the mean value and standard deviation of the target image patch were calculated respectively. Then a sort queue and two adaptive thresholds were set to discard unsuitable patches with certain features that differ greatly from the target patch. It was realized that the candidate patch in the next frame was detected and calculated with priority when whose mean values and standard deviations were both within a certain margin of the target patch. Experimental results show that the proposed algorithm increases the frame rate about 10% than the original KCF algorithm, and its tracking accuracy is about 2.2%, 14.4% and 24.9% higher than that of other algorithms such as KCF, circulant structure of tracking-by-detection with Kernels and Struc, respectively.

**Keywords:** object tracking; correlation filter; candidate image patch; search strategy

目标跟踪是计算机视觉领域中的研究热点,在民用和军事领域都有着非常广泛的应用。民用方面包括视频监控、人机交互、虚拟现实、移动机器人等;而军事方面则包括无人机、战场监控、精确制导以及空中预警等。运动目标跟踪,即通过目标的有效表达,给定第一帧图像中目标初始状态(通常为位置或范围),在图像序列中寻找与目标模板最相似的候选目标区位置的过程<sup>[1]</sup>。尽管多年来目标跟踪领域已取得了不错的进展,但它

仍是一个富有挑战性的问题。跟踪技术的研究重点依旧在于:如何解决部分遮挡、平面外旋转、背景相似以及尺度改变等<sup>[2]</sup>。如何建立有效的外观模型是跟踪算法能否成功的关键。

最早将相关滤波用于目标跟踪的是误差最小平方和滤波器<sup>[3]</sup>(minimum output sum of squared error filter, MOSEF),它采用灰度特征,并用峰值旁瓣比来判断目标是否被遮挡或跟踪失败。核循环结构检测跟踪<sup>[4]</sup>(circulant structure of tracking-

收稿日期:2018-08-28

作者简介:杨志方,硕士,副教授。Email:wit\_chuangxin@163.com

引文格式:杨志方,陈曦. 优化搜索策略的KCF目标跟踪算法[J]. 武汉工程大学学报, 2019, 41(1): 98-102.

by-detection with kernels, CSK), 速度大幅领先其之前的滑窗检测法。而文献<sup>[5]</sup>引入了循环矩阵和核的概念,并在核相关滤波器中提取方向梯度直方图特征,利用循环矩阵在傅里叶空间可对角化的性质大大降低了运算量。高效卷积算子的跟踪<sup>[6]</sup>(efficient convolution operators for tracking, ECO),从模型大小、样本集大小和更新策略3个方面加速,在多个顶尖目标跟踪数据集上性能都是第一,而且没有过拟合的问题。文献<sup>[7]</sup>在判别相关滤波器的框架下引入一种空间正则化组件,根据空间位置惩罚相关滤波器系数,扩大了训练图像区域并提高了分类器性能。文献<sup>[8]</sup>计算采样后分块粒子的权值函数,再利用霍夫投票机制来估计目标位置。

1 核相关滤波跟踪算法

1.1 循环矩阵及移位过程

核化相关滤波器的高速跟踪 (high-speed tracking with kernelized correlation filters, KCF), 是目标跟踪领域一种经典的算法,具有很高的研究价值。该算法利用循环矩阵和离散傅里叶变换以及核方法解决了复杂的矩阵运算问题,提升了跟踪性能和相关滤波的鲁棒性<sup>[9]</sup>。

文献<sup>[5]</sup>利用岭回归去训练跟踪器。跟踪器在目标的中心位置将目标与周围的小部分背景选做跟踪窗口,得到大小  $M \times N$  的样本图像块  $x$ 。再对  $x_i$  进行循环移位,  $i \in \{0, 1, \dots, M-1\} \times \{0, 1, \dots, N-1\}$ , 而每一个  $y_i$  都对应于一个样本  $x_i$  的高斯标签(正0负1)。设训练样本集为  $(x_i, y_i)$ , 则其线性回归函数为  $f(x_i) = \omega^T \odot x_i$ 。  $\omega$  为列向量, 表权重系数。  $\omega$  通过最小二乘法求解:

$$\min_{\omega} \sum_i (f(x_i) - y_i)^2 + \lambda \|\omega\|^2 \tag{1}$$

其中  $\lambda$  是控制过拟合的正则化参数, 以保证分类器的泛化性能。  $\odot$  表示元素点乘。

该最小化目标函数在频域中有如下闭式解:

$$\omega = (X^H \cdot X + \lambda \cdot I)^{-1} X^H Y \tag{2}$$

其中,  $X^H$  为 Hermitian 转置, 即  $X^H = (X^*)^T$ ,  $X^*$  为  $X$  的复数共轭。

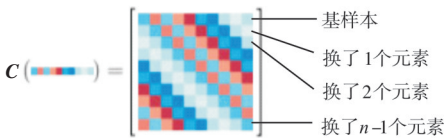


图1 一维向量得到的循环矩阵

Fig. 1 Cyclic matrix obtained from one-dimensional vector

对二维图像, 可通过循环移动  $x$  轴和  $y$  轴实现不同位置的移动。因此由一个向量  $x \in R^n$  可通过不断地乘上排列矩阵得到  $n$  个循环移位向量, 如图1所示。

再将这  $n$  个向量依序排列到一个矩阵中, 就形成了  $X$  生成的循环矩阵  $C(X)$ 。在傅氏空间中, 任意的  $C(X)$  都能使用矩阵  $F$  来进行对角化:

$$X = F \cdot \text{diag}(\hat{x}) \cdot F^H \tag{3}$$

其中  $F$  是离散傅里叶矩阵。  $\hat{x} = F(x)$  表示矩阵  $X$  的第一行向量经过离散傅里叶变换后的值。循环移位得到的样本如图2所示。

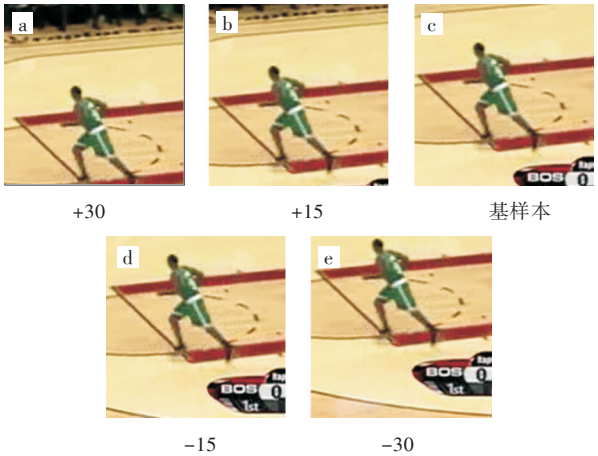


图2 二维图像经过不同行数的移位:

(a)下移30行, (b)下移15行, (c)基样本, (d)上移15行, (e)上移30行

Fig. 2 Two-dimensional images shifting through different lines: (a)down 30, (b)down 15, (c)sample, (d)up 15, (e)up 30

1.2 滤波器系数及相关响应的求解

利用核方法, 将线性问题的输入映射到非线性特征空间  $\Phi(x)$  中。过程分为如下两步:

1) 将解  $\omega$  表达为样本的线性组合:  $\omega = \sum_i \alpha_i \cdot \phi(x_i)$ 。其中  $\alpha_i$  为  $\omega$  的对偶空间向量, 即滤波器系数。

2) 用点积的形式来写入核方法:  $k(x, x') = \phi^T(x') \cdot \phi(x)$ , 该方法用核函数  $k$  (高斯或多项式) 来计算。所有样本对间的点积保存在一个  $n \times n$  的核矩阵  $K$  中, 元素值为  $K_{ij} = k(x_i, x_j)$ 。

但核方法的缺点在于, 回归函数的复杂性会随样本数目的增加而增加:

$$f(z) = \omega^T \odot \phi(z) = \sum_{i=1}^n \alpha_i \odot k(z, x_i) \tag{4}$$

存在如下定理: 给定循环数据  $C(x)$ , 对任意排列 (置换) 矩阵  $M$ , 当核函数满足  $k(x, x') =$

$k(\mathbf{M}x, \mathbf{M}x')$ , 则对应的  $\mathbf{K}$  为循环矩阵。而核化的岭回归的解为:  $\alpha = (\mathbf{K} + \lambda \cdot \mathbf{I})^{-1} \mathbf{y}$ 。其中  $\mathbf{K}$  为核矩阵,  $\alpha$  为系数  $\alpha_i$  的向量, 对偶空间中的解。再像线性情况一样对角化上述方程, 有:  $\hat{\alpha} = \frac{\hat{\mathbf{y}}}{\hat{\mathbf{k}}^{xx} + \lambda}$ 。其中  $\mathbf{k}^{xx}$  为核矩阵  $\mathbf{K} = \mathbf{C}(\mathbf{k}^{xx})$  的第一行的向量,  $\hat{\mathbf{y}}$  为标签列向量的傅里叶变换。

很少单独对一个图像块来评估回归函数  $f(z)$ 。为快速检测感兴趣物体, 一般在若干图像位置上评估  $f(z)$ , 这些块可用循环移位来建模<sup>[10]</sup>。再用  $\mathbf{K}^z$  表示所有训练样本与候选块间的核矩阵。而所有训练样本和图像块都分别为基样本  $x$  和基块  $z$  的循环移位。用循环矩阵的第一行向量来定义核矩阵:  $\mathbf{K}^z = \mathbf{C}(\mathbf{k}^z)$ , 由方程(4)可计算所有候选块的回归函数:  $f(z) = (\mathbf{K}^z)^T \cdot \alpha$ 。其中  $f(z)$  为检测响应, 包含了所有  $z$  循环移位的输出。

对于高斯核, 它是欧氏距离  $\|x_i - x_j\|^2$  的函数, 即  $\|x_i - x_j\|^2 = \|x_i\|^2 + \|x_j\|^2 - 2x_i^T \cdot x_j$ 。则有

$$\mathbf{K}^{xz} = h(\|\mathbf{x}\|^2 + \|\mathbf{z}\|^2 - 2\mathbf{C}(\mathbf{x}) \cdot \mathbf{z})^T = \exp(-\frac{1}{\sigma^2}(\|\mathbf{x}\|^2 + \|\mathbf{z}\|^2 - 2\mathbf{F}^{-1}(\hat{\mathbf{x}}^* \odot \hat{\mathbf{z}}))^T)$$

为了有效计算上述回归函数, 结合上式并对角化  $f(z)$ , 有

$$\hat{f}(z) = \hat{\mathbf{k}}^{xz} \odot \hat{\alpha} \quad (5)$$

## 2 优化搜索策略的改进算法

如1.1节所述, 分类器在  $M \times N$  的样本图像块  $x$  上训练, 并以目标位置为每一块的中心。当训练结束后, 对于后续帧, 同样以上一帧目标位置为中心(预测区域), 在大小为  $M \times N$  的图像块  $z$  (patch) 上检测并计算响应。检测的候选区域便是由上一帧的预测区域做循环移位得到的, 如图3所示。通过密集采样可以提取出图像块的全部特征, 优于只利用每个候选框的局部特征的随机采样。

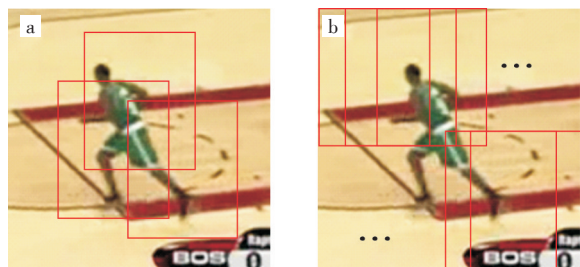


图3 检测图像块:(a)随机采样,(b)密集采样

Fig. 3 Sampling of detected image patches:

(a) random sampling, (b) dense sampling

虽然密集采样可以利用图像块的全部特征, 但是信息冗余非常大, 并且会降低搜索速度。由实验观察到, 视频前后两帧中的两个目标区域块, 其均值和方差应该最为接近。基于此提出了一种优化搜索策略, 用于改进 KCF 算法的密集采样部分, 以达到提高搜索速度, 实现目标跟踪实时性的目的。

在第  $t$  帧中, 由最大响应得到了目标的位置, 该图像块大小为目标框加扩大 1.5 倍的 padding 窗口。其数据存放在二维数组中, 数组的元素即为图像的像素。先将第  $t$  帧中图像块的均值和标准差记为  $\mu_m$  和  $\sigma_m$ 。而下一帧中, 经循环移位的  $n$  个图像块的均值和标准差分别为:  $\mu_1, \mu_2, \dots, \mu_n$  和  $\sigma_1, \sigma_2, \dots, \sigma_n$ 。再它们与  $\mu_m$  和  $\sigma_m$  作比较, 满足以下条件的图像块优先检测并计算响应:

$$\begin{cases} \mu_m - \mu_i < T_1, & i = 0, 1, \dots, n \\ \sigma_m - \sigma_i < T_2, & i = 0, 1, \dots, n \end{cases} \quad (6)$$

手动选择阈值并测试数据集集中的 6 个视频序列 (Basketball, Bird2, Box, Car4, Freeman1, Girl), 本文算法的平均跟踪帧率提升为 10.1%~12.1%。然而, 在实际跟踪过程中, 多次手动设置阈值会影响跟踪效率, 统一设置固定阈值又会降低算法的自适应性, 使得跟踪准确度受到影响<sup>[11]</sup>。

为避免上述问题, 本文将阈值设为能够使算法的平均跟踪帧率提升 10% 时的阈值大小。为求得该自适应阈值, 需要先确定平均每帧所需的用于匹配的图像块个数。假设算法帧率提升达到 10% 时, 平均每帧需要用到  $k$  ( $k < n$ ) 个候选图像块用来跟踪并计算目标的中心位置。再依据最相似图像块间的均值和标准差的距离应最小的原理, 这  $k$  个候选块取自排序队列中距离最小的前  $k$  个, 其中, 排序队列按照如下公式计算所得的距离大小进行排列:

$$d_i = \alpha |\mu_m - \mu_i| + (1 - \alpha) |\sigma_m - \sigma_i| \quad i \in (1, n) \quad (7)$$

公式(7)中,  $d_i$  反应了当前帧中的图像块  $i$  与上一帧中的目标图像块  $m$  之间的均值和标准差的距离大小,  $\alpha \in (0, 1)$  为排序时分别赋予均值和方差的权重, 当图像块的均值对跟踪结果的影响更大时, 算法取  $\alpha > 0.5$ , 当图像块的标准差对跟踪结果影响更大时, 算法取  $\alpha < 0.5$ 。将全部  $n$  个块都按上述公式算出值并按由小到大的顺序排序, 那么, 序列中的前  $k$  个距离所对应的  $k$  个图像块, 即为算法在实际跟踪时用于匹配的  $k$  个图像块。后



续 $n-k$ 个块由于误差较大,因此舍弃不用。

设定平均帧率提升10%的情况下,可以确定每帧平均所需的匹配图像块个数 $k$ ,在此基础上,阈值 $T_1, T_2$ 可以通过上一帧目标中心图像块 $m$ 与当前帧第 $k$ 块之间的均值、标准差的差值得到。而且每个测试集由于图像的像素不同(720×400, 576×432等)和面临的问题不同(比如遮挡、背景相似、尺度变化等),阈值会根据满足标准的 $\mu_k、\sigma_k$ 自动调整,达到自适应的效果。

3 结果与讨论

实验平台为 Windows7 64 位旗舰版系统和 MATLAB R2014a 软件,在笔记本电脑上进行了调试运行。CPU 为英特尔酷睿 i5-3210M @ 2.50 GHz 双核,内存为 DDR3 1 600 MHz 4 GB<sup>[12]</sup>。

选取 OTB100<sup>[13]</sup> 上的 6 个具有挑战性的视频序列 (Basketball, Bird2, Box, Car4, Freeman1, Girl) 进行测试,并对比分析了实验结果。这些序列中的目标会受到诸如光照变化、快速运动、尺度变化、遮挡以及平面外旋转等多种复杂因素的影响。实验中改进算法的 MATLAB 源码是由作者公开的个人网站上获取的,可以保证实验数据的公正客观。

3.1 候选图像块数量的影响

实验首先测试了算法帧率是否会受到图像块数量的影响。将原文中用于扩展背景信息的 padding 窗口由 1.5 倍目标框扩大倍数修改为 0.5 和 1.0,再进行对目标候选区域的分块测试。实验结果表明改变 padding 大小对跟踪的精度并无显著影响,在 3 个视频序列上分为不同大小图像块的帧率如表 1 所示。

表 1 不同 padding 大小下的算法帧率

Tab. 1 Algorithm frame per second performance at different padding sizes frames/s

视频序列名称	Padding 填充倍数		
	0.5	1.0	1.5
篮球	108.02	109.24	112.61
盒子	14.84	21.94	33.40
女孩	147.35	148.49	150.32

从表 1 可得出 padding 越小时算法帧率也越小的结论。由于 padding 较小时会增加预测区域循环移位的次数,导致分块得到的 patch 数量增加,使公式(5)计算核相关的次数也相应地增加,因此算法的速度会稍微降低。本实验未对 padding 做更大值的测试,原因是分块尺度过大,搜索窗可能

占满整个图像甚至超过原图像的尺寸,故无法对原图进行分块,也无法对目标进行跟踪。因此, padding 为 1.5 倍大小时跟踪效果最显著。

3.2 不同算法对比分析

实验中控制过拟合的正则化参数 $\lambda$ 设为 $10^{-4}$ 。高斯核条件下的 $\sigma$ 为 0.2。padding 窗经过上述分析取 1.5。 $k$ 在 Basketball 序列中取 13(总共 40 个),即前 32.5% 的图像块,在 Girl 序列中取 11(总共 24 个),即前 45.8% 的图像块。自适应阈值 $T_1$ 和 $T_2$ 根据上述不同 $k$ 值计算出的 $\mu_k、\sigma_k$ 自动调整时效果较好。

在 MATLAB 上选用距离精度作为衡量算法性能的指标,目标的中心位置误差(center position error, CLE)小于某一固定阈值时,符合的帧数与视频总帧数的比值即为算法精度。这里的阈值通常以 20 个像素返回精度。

图 4 的实验结果集成了改进算法 OSKCF (Optimized Searching Strategy for KCF, OSKCF) 与较早的 9 种性能较好的跟踪算法,并横向对比了其中 CSK、KCF、Struck<sup>[14]</sup> (structured output tracking with kernels, Struct) 和检测学习跟踪<sup>[15]</sup> (tracking-learning-detection, TLD) 4 种算法的性能。横坐标为目标误差像素阈值,以 10 像素为 1 个单位,纵坐标则表示算法精度。各算法曲线基本呈半梯形(左半部分)状,阈值在 30 像素后曲线基本不再增长,因此取 20 像素来计算阈值是合理的。

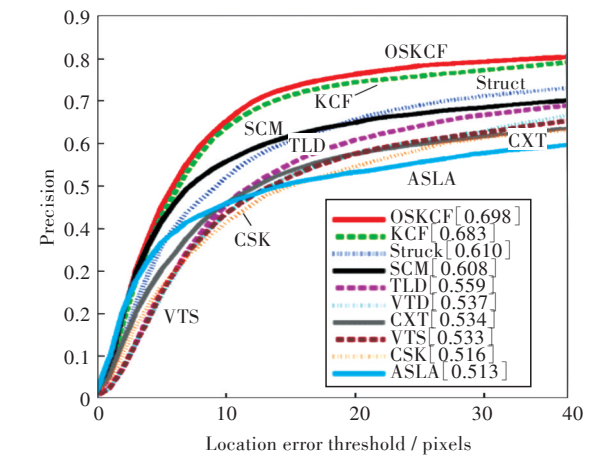


图 4 不同跟踪算法的精度图

Fig. 4 Precision of different tracking algorithms

从图 4 可看出,经过优化候选区域的搜索策略并筛选图像块后,改进算法 OSKCF 的性能在较早的 10 个测试算法中最好,优于 2012 年以前最快的算法 Struct 以及经典的 long-term 算法 TLD。在测试 Basketball 序列时,OSKCF 平均精度为 0.942,平

均帧率为 115.83 帧/秒。而原 KCF 平均精度和帧率为 0.923 和 103.34 帧/秒,帧率提升达 12.1%;测试 Girl 序列时,OSKCF 平均精度和帧率分别为 0.879 和 157.67 帧/秒,原 KCF 分别为 0.864 和 143.25 帧/秒,帧率提升 10.1%。而测试其他几个挑战性更大的序列,涉及目标的快速运动及遮挡问题,精度较低只有 50%左右。本算法的精度较 KCF, Struct, TLD, CSK 分别提升 2.2%, 14.4%, 24.9%, 35.3%。测试 6 个视频序列的平均帧率如表 2 所示。

表 2 算法在 6 个视频序列上的平均帧率  
Tab. 2 Average frame rates of algorithms in  
6 video sequences frames/s

算法	视频序列名称					
	对比	篮球	鸟	盒子	轿车	市民 女孩
原 KCF		103.34	51.56	30.14	34..45	257.03 143.25
改进 KCF		115.83	56.20	32.55	38.24	284.79 157.67

经过分析,优化搜索策略后的 OSKCF 虽然对跟踪精度提升不大,但是可以明显提升原算法的帧率,从而使得目标跟踪算法的实时性进一步得到提高。

4 结 语

本文提出了一种优化候选图像块搜索策略的 OSKCF 算法,通过  $T_1$  和  $T_2$  这两个自适应阈值和最相似图像块间的均值方差的距离应最小的原理,选取排序队列中距离最小的前  $k$  个候选图像块,明显提升算法的 FPS。实验结果表明,本文算法较原算法速度提升 10%左右,性能较好,可满足大多数情况下的目标跟踪检测需求。但是在目标受到诸如光照变化、快速运动、遮挡以及平面外旋转等多种复杂因素的影响时,部分测试视频出现无法准确跟踪目标的现象。如何解决这些当前跟踪领域的一些共同难点,避免产生跟踪漂移等问题,将在未来的工作中做进一步研究。

参考文献

[1] 张微,康宝生. 相关滤波目标跟踪进展综述[J]. 中国图象图形学报, 2017,22(8):1017-1033.

[2] 尹宏鹏,陈波,柴毅,等. 基于视觉的目标检测与跟踪综述[J]. 自动化学报, 2016,42(10):1466-1489.

[3] BOLME D S, BEVERIDGE J R, DRAPER B A, et al. Visual object tracking using adaptive correlation filters [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco: IEEE,

2010:2544-2550.

[4] HENRIQUES J F, CASEIRO R, MARTINS P, et al. Exploiting the circulant structure of tracking-by-detection with kernels [C]// Proceedings of 2012 European Conference on Computer Vision. Heidelberg Berlin:Springer, 2012:702-715.

[5] 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.

[6] DANELLJAN M, BHAT G, KHAN F S, et al. ECO: efficient convolution operators for tracking [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017: 6931-6939.

[7] DANELLJAN M, BHAT G, KHAN F S, et al. Learning spatially regularized correlation filters for visual tracking [C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2016: 4310-4318.

[8] LI Y, ZHU J K, HOI S C H. Reliable patch trackers: robust visual tracking by exploiting reliable patches[C]// Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 353-361.

[9] XU Y L, WANG J B, LI H, et al. Patch-based scale calculation for real-time visual tracking[J]. IEEE Signal Processing Letters, 2016, 23(1):40-44.

[10] 孙健,向伟,谭舒昆,等. 改进的核相关滤波跟踪算法[J]. 计算机工程与应用, 2018,54(9):178-182.

[11] 单玲玉,闵锋,李延达. 全局阈值与局部阈值相结合的视网膜血管分割方法[J]. 武汉工程大学学报, 2015,37(3):62-67.

[12] 张雷,王延杰,孙宏海,等. 采用核相关滤波器的自适应尺度目标跟踪[J]. 光学精密工程, 2016,24(2): 448-459.

[13] WU Y, LIM J, YANG M H. Online object tracking: a benchmark [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013:2411-2418.

[14] HARE S, SAFFARI A, TORR P H S. Struct: structured output tracking with kernels [C]// Proceedings of 2011 IEEE International Conference on Computer Vision. Barcelona:IEEE, 2012:263-270.

[15] KALAL Z, MATAS J, MIKOLAJCZYK K. Tracking-learning-detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34 (7) : 1409-1422.