

融合多层深度特征的核相关滤波跟踪算法^①李国友^{②*} 纪执安^{③*} 张凤煦^{**}

(* 燕山大学电气工程学院 秦皇岛 066004)

摘 要 针对核相关滤波算法(KCF)难以处理目标尺度变化、旋转、遮挡等问题,本文在 KCF 的框架下提出了一种融合多层深度特征的抗遮挡目标跟踪算法。首先,在频域中,利用岭回归分类器训练 VGG-2048 上的 conv3 和 conv6 两层深度特征,分别得出置信度,将两者特征加权相连,替代原 KCF 的方形梯度直方图(HOG)特征,同时引入第 1 帧目标的残留信息,获得更为出色的位置响应输出。然后,针对遮挡问题,提出一种响应峰值判断抗遮挡机制。最后,通过双线性插值建立深度特征尺度池,解决目标尺度问题。在测试集(OTB-100)上的实验结果表明,改进后的算法能解决复杂环境下的目标跟踪问题,算法具有鲁棒性。

关键词 核相关滤波算法(KCF);深度特征;响应峰值;尺度池;鲁棒性

0 引言

目标跟踪技术是计算机视觉的重要分支,随着科学技术的发展,其在多个领域广泛应用^[1],如智能视频监控、人机互动、手势识别、虚拟现实等^[2]。通常目标跟踪算法根据第 1 帧目标框的信息,跟踪整个视频序列感兴趣的区域。但在实际应用中,由于光照、目标旋转、遮挡等干扰^[3],跟踪过程中会产生漂移,甚至会跟丢标定目标。因此,设计一个鲁棒的目标跟踪算法是众多学者追求的热门课题。

传统的跟踪算法主要是对目标建立有效决策模型和提取有用的外观特征,提取的目标特征多为灰度、颜色、方形梯度直方图(histogram of oriented gradient, HOG)特征等^[4],统称为人工提取特征。相关滤波目标跟踪算法,如最小输出和的平方误差滤波跟踪^[5](minimum output sum of squared error filter, MOSSE)、循环结构核的目标跟踪^[6](circulant structure kernel, CSK)均是先对目标进行提取人工特征,

然后通过分类器进行训练,最后完成对目标的跟踪。在相关滤波算法中,最为出色的就是 Henriques 等人^[7]提出的核相关滤波算法(kernel correlation filter, KCF),KCF 通过循环矩阵和岭回归分类器^[8]训练一个目标检测器,来判断目标的位置。

但人工特征仅仅能直观获得目标的外观信息,不能很好捕捉目标的语义信息^[9],对复杂环境下的目标跟踪不具有鲁棒性。最近,卷积神经网络^[10](convolutional neural network, CNN)在计算机视觉领域有了突破性进展,其能对图像信息进行更深层次的数据挖掘,主要应用于图像分割、动作识别^[11]、图像分类。在目标跟踪中,基于 CNN 的深度特征跟踪算法也相继提出。文献[12]中, Danelljan 等人很早将卷积特征应用在相关滤波器框架中,和标准人工特征相比,卷积特征提供了更好的跟踪效果。Mozhdehi 等人^[13]结合卷积特征图,通过粒子滤波器估计标定目标的大小和位置。Zhang 等人^[14]在自表达相关估计方法的基础上,提出了一种新的主动半监督学习方法。尽管现有的 CNN 跟踪器实现了最先进

① 国家自然科学基金(F2012203111)资助项目。

② 男,1972 年生,博士,教授;研究方向:工业控制,图像处理,机器视觉,智能控制;E-mail: lgyysu@163.com

③ 通信作者,E-mail: 985003629@qq.com

(收稿日期:2019-06-15)

的性能,但仍然有一些局限性,这些算法中的大多数只使用 CNN 的单层卷积特性来表示目标对象。因此,为了获得更好的跟踪性能,必须将不同卷积层的特性结合起来,以最优方式表示前景对象,并将其与背景杂波分离。

本文受 KCF 算法和深度特征的启发,基于 KCF 框架,提出了以下 3 种改进。(1)利用 ImageNet 上训练好的 VGG-2048 2 个卷积层(conv3 和 conv6),学习其中深度特征,其后用岭回归分类器分别训练不同卷积层的深度特征,依次得出置信度得分,接着依据得分高低将 2 种特征加权相连,实现特征的重复利用,用融合后的多层深度特征代替 KCF 中的人工特征(HOG 特征)。同时引入了第 1 帧目标的丰富信息,得到最终的位置输出响应。(2)对于目标跟踪中的全遮挡问题,设计一个抗遮挡机制,通过波峰震荡程度判断目标遮挡,当视频相邻 2 帧的峰值

小于给定阈值时,停止对模型更新,最终实现对跟踪目标的精准定位。(3)对于目标跟踪过程中的尺度变化,本文通过双线性差值建立了 6 个深度特征尺度池,利用分类器进行训练,得到的最大响应即为目标的最佳尺度。

1 KCF 算法分析

核相关滤波算法(KCF)是一种鉴别式跟踪算法,其主要创新是通过循环矩阵结构提高目标跟踪速度和识别能力^[15]。如图 1 所示,算法主要流程如下:确定跟踪区域后提取其 HOG 特征,在频域中,对每个特征通道进行加权余弦窗口操作^[16]。引入多项核函数匹配目标区域和待选择区域,选取响应峰值最高的得分即为跟踪目标。

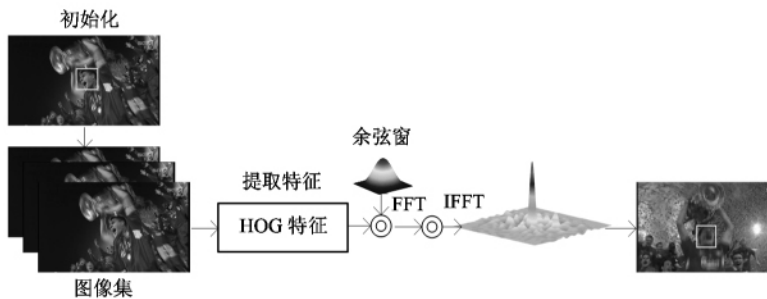


图 1 KCF 流程框图

1.1 岭回归分类器训练

设训练集样本 (x_i, y_i) , 其线性函数表示为 $f(x_i) = \mathbf{w}^T x_i$, 训练的目的在于通过样本 x_i 找到一个分类器 $\mathbf{w}$ 让回归目标 y 最小, 让岭回归发挥最大性能, 其中最小输出误差公式为

$$\min_{\mathbf{w}} \sum_i (f(x_i) - y_i)^2 + \lambda \|\mathbf{w}\|^2 \quad (1)$$

其中 λ 用于控制系统的复杂性, 保证分类器的泛化性能, 后导出分类器 $\mathbf{w}$:

$$\mathbf{w} = (\mathbf{X}^H \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^H \mathbf{y} \quad (2)$$

其中, $\mathbf{X}$ 为样本矩阵, $\mathbf{y}$ 为回归目标矩阵, $\mathbf{I}$ 为单位矩阵。

上述方程为线性计算, 可降低存储和计算复杂性数量等级来提高跟踪器速度。对于非线性计算, 引入映射函数 $\varphi(x)$, 使原本线性回归转化为非线

性回归问题, 得出:

$$\mathbf{w} = \sum_i \alpha_i \varphi(x_i) \quad (3)$$

利用循环矩阵和离散傅里叶变化, 可得核化的最小二乘解为

$$\hat{\alpha} = \frac{\hat{\mathbf{y}}}{\mathbf{k} + \lambda} \quad (4)$$

式(4)中核函数 k 可用下式表示:

$$k = \exp\left(-\frac{1}{\sigma^2}(\|x\|^2 + \|x'\|^2 - 2F^{-1}(\hat{x} \cdot \hat{x}'))\right) \quad (5)$$

其中 σ 为带宽, x 为检测样本。

1.2 快速检测

每次更新目标的位置, 岭回归分类器 $\mathbf{w}$ 以 $m \times n$ 的窗口来搜索候选区域 z , 其为目标的概率分布公

式为

$$f(z) = F^{-1}(\hat{a} \cdot \hat{k}_{xz}) \tag{6}$$

其中 $f(z)$ 表示包含 z 的所有循环位移输出向量, k_{xz} 表示 x 和 z 的核相关, $f(z)$ 得到的最大值即为跟踪目标。

1.3 模型更新

为满足目标过程中的变化,需要对外观模型 x 和分类器系数 a 实时更新,更新公式如下:

$$\begin{aligned} x_i &= (1 - \eta)x_{i-1} + \eta x_i \\ a_i &= (1 - \eta)a_{i-1} + \eta a_i \end{aligned} \tag{7}$$

其中, x_i 和 x_{i-1} 分别为当前帧和上一帧的目标特征模型, a_i 和 a_{i-1} 是当前帧和上一帧的系数矩阵, η 为学习系数。

2 融合多层深度特征的抗遮挡 KCF 算法

在目标跟踪算法中,相比于传统的手工特征

(如颜色、形状、梯度)提取,深度特征具有很好的学习能力。但单层卷积网络提取的深度特征不能处理目标本身发生的巨大变化,在恶劣环境下会产生响应峰值偏移,进而跟丢目标。为了得到期望的峰值响应和目标准确位置,本文基于 KCF 框架,利用低层卷积特征(conv3)的位置信息和高层卷积层特征(conv6)的语义信息,将 2 种特征进行加权处理后相连,实现特征的重复利用。同时,引入了第 1 帧目标的残留信息并加入了尺度池,最终获得目标的位置和尺度,图 2 为本文算法总体思路图。

2.1 融合响应输出

鉴于 CNN 中进行池化操作会影响目标特征图的尺寸,更深层次的卷积层,会降低空间分辨率,缩小目标的尺度,导致无法对目标进行准确定位。因此,可利用双线性插值法^[17]调整目标特征图的大小:

$$x_i = \sum_k \tau_{ik} h_k \tag{8}$$

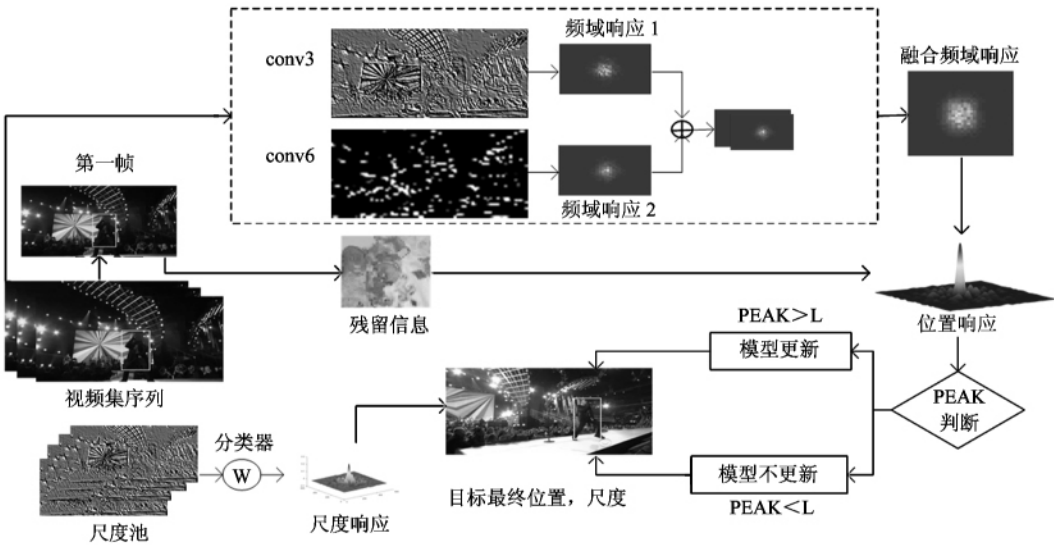


图 2 融合多层深度特征的抗遮挡 KCF 算法框图

其中, h 是池化操作前的初始特征图, τ 由相邻两特征向量 i 和 k 的位置决定, x 是经双线性插值处理后的特征图。由 2.2 节中的公式可知,单层深度特征在频域中的响应为

$$f_i = F^{-1}(\hat{a} \cdot \hat{k}_{xz}) \tag{9}$$

在特征融合的过程中,随着卷积层的增加,深度

特征会赋予不同的权重系数,越深层的特征往往会得到更大的置信度。低层卷积特征(conv3)和高层卷积层特征(conv6)响应叠加为

$$\Omega = \sum_{i=1}^2 w_i f_i \tag{10}$$

其中, Ω 为融合双层深度特征的响应, w_i 为不同卷积层特征的权重系数。

由于目标跟踪过程中,第1帧的目标信息往往最丰富,因此,引入了第1帧中的目标残留信息,能得到最出色的响应输出:

$$\psi = \sum_{i=1}^2 w_i f_i + f_1(x_1) \quad (11)$$

其中, ψ 为加入第1帧信息的响应输出, x_1 是第1帧目标的深度特征, f_1 为第1帧目标的频域响应。

2.2 抗遮挡处理

针对目标追踪过程中半遮挡、全遮挡等干扰问题,本文引入一个遮挡判断机制。如图3所示,在KCF跟踪中,每个跟踪目标对应一个高斯标签响应。在目标未遮挡时,响应图震荡平缓,仅出现一个峰值,在目标遮挡时响应图震荡剧烈,出现多峰值。可以根据峰值来判断指定跟踪目标是否被遮挡。

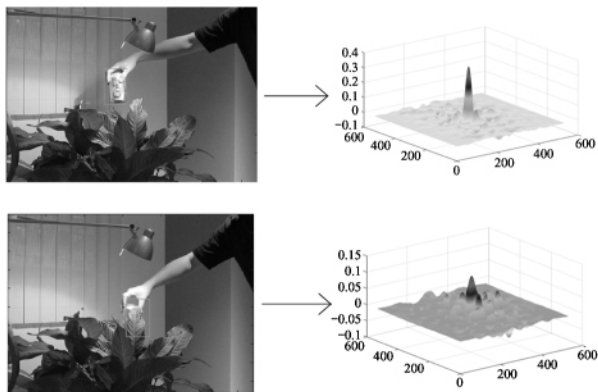


图3 有无遮挡波峰震荡图

根据文献[18],得出下式计算波峰震荡程度:

$$APEK = \frac{|f_{\max} - f_{\min}|^2}{\text{mean}((\sum_{w,h} f_{w,h} - f_{\min})^2)} \quad (12)$$

其中, $f_{\max}$ 表示响应最高的位置, $f_{\min}$ 表示响应最低的位置, $f_{w,h}$ 为 (w, h) 位置上的响应, L 为遮挡阈值。当视频序列相邻两帧之间的 $APEK$ 差小于遮挡阈值时,表示目标被遮挡,停止模型的更新。引入了峰值判断,最后目标的位置 $(\tilde{x}, \tilde{y})$ 由 ψ 的最大响应决定。

$$(\tilde{x}, \tilde{y}) = \underset{x, y}{\operatorname{argmax}} \psi(x, y) \quad (13)$$

2.3 尺度处理

在实际跟踪过程中,外界环境的干扰会使目标的尺度发生变化。KCF的跟踪框是固定的,其不会随着目标的大小而改变。为此在每个卷积层下,本文通过双线性插值调整深度特征图大小 $M \times N$, 建

立从小到大6个尺度作为尺度池 $S = \{s_1, s_2, s_3, \dots, s_6\}$, 将每个尺度通过岭回归分类器训练,得到最大的响应即为目标尺度:

$$v = \operatorname{argmax}_s \{ \max_j f(y_{ij}) \} \quad (14)$$

其中, i 为卷积层数, j 代表目标从低到高的尺度数, y 为频域中的尺度响应。

3 实验分析

3.1 实验环境设置

本实验的硬件环境为内存为8GB, Inter(R) Core(TM) i5-44600(3.20Hhz)的CPU台式机,使用了Matlab2015b作为软件平台进行仿真。实验参数设置为:遮挡阈值 $L = 27$, 第3层和第6层的卷积特征权重系数 $w = \{0.8, 0.5\}$, 规则化范数 $\lambda = 1 \times 10^{-4}$ 。

3.2 定量评价

为了验证本文改进的算法,主要利用中心位置误差和重叠率2个计算公式。中心位置误差公式定义为

$$CLE = \sqrt{(x_i - x_s)^2 + (y_i - y_s)^2} \quad (15)$$

其中, (x_i, y_i) 表示实际跟踪的目标位置, (x_s, y_s) 表示标准跟踪的目标中心位置, CLE 值越小,表示跟踪效果越好。

重叠率公式定义为

$$OR = \frac{\text{area}(ROI_T \cap ROI_G)}{\text{area}(ROI_T \cup ROI_G)} \quad (16)$$

其中, ROI_T 为实际的跟踪窗口, ROI_G 表示标准的跟踪窗口。 OR 值越大,表示跟踪越佳。

实验中采用的测试集为 OTB-2015 中的 100 个标准集视频序列,所对比的算法为 KCF、CN^[19]、fDSST^[20]、Staple^[21] 等 35 种跟踪算法,实验指标为 OTB-2015 中的精确率图和成功率图。精确率图是中心位置误差(CLE)小于指定阈值的帧数占视频序列总帧数的百分比,本次阈值设定为 20 像素。成功率图指的是重叠率(OR)值大于阈值的帧数占视频序列总帧数的百分比,本次阈值设为 0.5。表1为6种跟踪算法在 OTB-2015 测试集中选取的7个具体视频序列的成功率数值,从表中看出,除了 Jogging

和 CarScale 视频序列外,本文提出的算法(OWN)在其他 5 个视频序列(MotorRolling、David、CarScale、Couple、Coke)中,平均跟踪成功率最大,整体平均成

功率为 0.85,相对于其他对比的跟踪算法,为最大值,对目标跟踪效果出色。

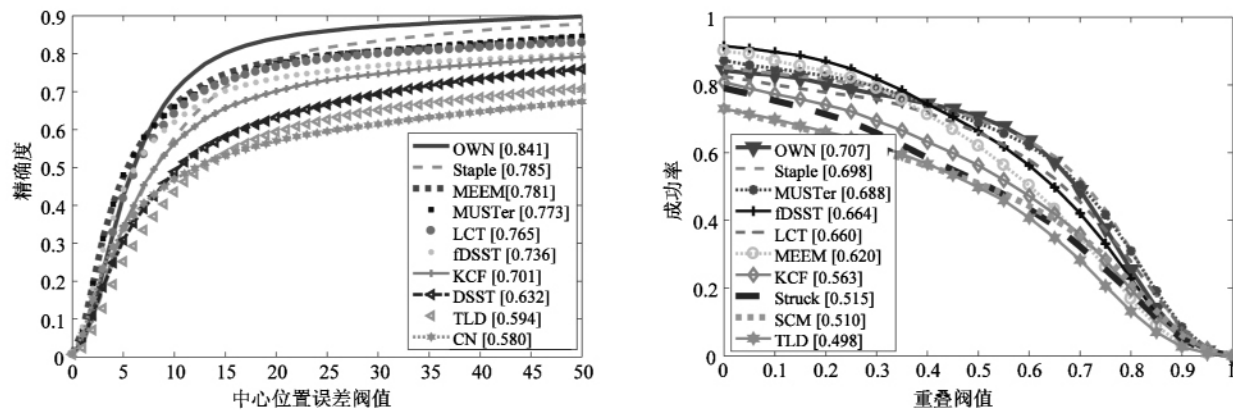


图 4 OPE 精确度图和成功率图

表 1 跟踪算法平均成功率

	KCF	fDSST	STC	TLD	MEEM	OWN
Jogging	0.29	0.35	0.31	0.78	0.85	0.81
MotorRolling	0.22	0.31	0.32	0.26	0.29	0.76
David	0.75	0.79	0.32	0.51	0.85	0.87
CarScale	0.77	0.88	0.51	0.54	0.81	0.83
Couple	0.71	0.61	0.27	0.33	0.86	0.90
Coke	0.29	0.18	0.41	0.44	0.82	0.86
Basketball	0.91	0.90	0.43	0.52	0.93	0.95
Average	0.56	0.57	0.37	0.48	0.77	0.85

如图 4 所示,本次实验仅显示排名前 10 的跟踪算法。粗实线(OWN)代表融合多层深度特征的 KCF 算法,从精确率图看出,本文改进算法其精确率达到了 83.8%,相比原 KCF 提高了 16.8%。从成功率图看出,改进算法成功率为 70.2%,比原

KCF 算法提高了 21.4%。在成功率和精确率 2 个指标中,改进的 KCF 算法相比于其他 35 种跟踪算法,均排名第 1。表 2 为 8 种跟踪算法的跟踪速度,可看出 KCF 算法速度最快,跟踪实时性最好,本文算法虽然跟踪精度高,但 fps 仅为 2.5。

表 2 8 种跟踪算法速度比较

算法	fDSST	Staple	KCF	DSST	LCT	Struck	MEEM	OWN
速度(fps)	56.2	47.2	240.5	48.6	10.2	20.5	15.6	2.5

3.3 定性评价

本文了选取 Benchmark 网站上的 MotorRolling、David、Jogging、Carscale 4 个视频序列进行实验,如图 5 所示,这些视频包括多种复杂环境,如目标旋转、光照、遮挡、尺度变化等干扰条件。并与 DFT、

KCF、fDSST、TLD、Staple 5 种现有的跟踪算法进行对比。以下是不同场景下不同跟踪算法的实际效果分析。

(1) 快速运动及旋转

在 MotorRolling 视频中,所跟踪的摩托车进行

了旋转高速运动,目标轮廓发生变化,第 36 帧时,只有本文算法精确跟踪,其余算法均对目标发生漂移。在第 71 帧时,改进的 KCF 算法对目标发生局部漂移,但整体上能稳定跟踪摩托车目标。

(2) 尺度变化

在 David 和 CarScale 序列中,涉及目标尺度问题。David 中,目标所处的环境由暗到亮,6 种目标跟踪算法均能跟踪目标脸部特征,本文算法虽然大致能随指定目标进行尺度变化,但尺度上表现最为出色的是 fDSST 算法。在 CarScale 中,138 帧前,所

对比跟踪算法均能稳定跟踪目标,在第 252 帧时, TLD 算法跟丢目标, KCF 算法没有发生尺度变化,其余算法虽然对跟踪车辆进行了尺度变化处理,但都表现得太不出色,跟踪框没有和车辆大小保持一致。

(3) 遮挡物干扰

在 Jogging 视频序列中,目标在 71 帧时遇到了遮挡物,在第 109 帧时,目标离开遮挡物时,只有本文改进跟踪算法、TLD 算法、Staple 算法能继续跟踪目标,其余跟踪算法发生不同程度的漂移现象。



图 5 不同跟踪算法比较图

4 结 论

为解决目标在复杂环境下的跟踪问题,本文在 KCF 的基础上融合了多层深度特征,在频域中,将两者深度特征加权相连,通过岭回归分类器训练,同

时,引入了第 1 帧目标的丰富信息,获得最终的响应位置。针对目标跟踪中的遮挡干扰,利用响应峰值变化,当物体受到遮挡时,响应峰值会出现抖动多峰值的情况。因此,根据相邻两帧之间的峰值变化,来判断目标是否受到障碍物的干扰,若有遮挡物,则停

止对模型的更新。对于跟踪中的尺度变化问题,通过双线性插值提取不同大小的深度特征,建立尺度池,得到目标最佳尺度。最后将改进的 KCF 算法在 OTB-100 上进行测试,证明本文改进算法能处理跟踪过程的旋转、遮挡、尺度变化等问题,但跟踪算法实时性还有待进一步改进。

参考文献

- [1] 惠斌, 陈法领, 罗海波. 基于互信息的目标跟踪算法 [J]. 红外与激光工程, 2007, 36(5): 251-256
- [2] 李天成, 范红旗, 孙树栋. 粒子滤波理论、方法及其在多目标跟踪中的应用 [J]. 自动化学报, 2015, 41(12): 1981-2002
- [3] 王科平, 武帅帅, 王红旗. 自适应模型更新相关滤波目标跟踪方法 [J]. 传感器与微系统, 2019, 38(1): 1533-1539
- [4] 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, Amsterdam, Netherlands, 2016: 472-488
- [5] Bolme D S, Beveridge J R, Draper B A, et al. Visual object tracking using adaptive correlation filters [C] // Conference on Computer Vision and Pattern Recognition, California, USA, 2010: 13-18
- [6] 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, Florence, Italy, 2012: 1090-1097
- [7] Henriques J F, Caseiro R, Martins P, et al. High-speed tracking with kernelized correlation filters [J]. *Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 583-596
- [8] 赵东, 周慧鑫, 秦翰林, 等. 基于引导滤波和核相关滤波的红外弱小目标跟踪 [J]. 光学学报, 2018, 38(2): 2533-2540
- [9] Zhu X B, Zhang X M, Zhang X Y, et al. A novel framework for semantic segmentation with generative adversarial network [J]. *Journal of Visual Communication and Image Representation*, 2019, 58(12): 532-543
- [10] 李彦冬, 郝宗波, 雷航. 卷积神经网络研究综述 [J]. 计算机应用, 2016, 36(9): 2508-2515
- [11] Zhang X Y, Shi H, Li C, et al. Learning transferable self-attentive representations for action recognition in untrimmed videos with weak supervision [C] // Conference on Artificial Intelligence (AAAI), Hawaii, USA, 2019: 1-8
- [12] Danelljan M, Hager G, Khan F S, et al. Convolutional features for correlation filter based visual tracking [C] // 2015 IEEE International Conference on Computer Vision Workshop (ICCVW), Santiago, Chile, 2015: 3011-3018
- [13] Reza A, Yevgeniy M, Abubakar B, et al. Deep convolutional particle filter with adaptive correlation maps for visual tracking [C] // IEEE International Conference on Image Processing, Athens, Greece, 2018: 798-802
- [14] Zhang X Y, Shi H C, Zhu X B, et al. Active semi-supervised learning based on self-expressive correlation with generative adversarial networks [J]. *Neurocomputing (NEUCOM)*, 2019, 345(15): 103-113
- [15] Du B, Sun Y, Cai S, et al. Object tracking in satellite videos by fusing the kernel correlation filter and the three-frame-difference algorithm [J]. *Geoscience and Remote Sensing Letters*, 2017, 35(6): 1-5
- [16] Wang X, Brien M, Xiang C, et al. Real-time visual tracking via robust kernelized correlation filter [C] // IEEE International Conference on Robotics & Automation, Marina Bay Sands, Singapore, 2017: 3820-3827
- [17] 刘卜, 屈有山, 冯桂兰, 等. 小波双线性插值迭代算法应用于光学遥感图像 [J]. 光子学报, 2006, 35(3): 468-472
- [18] Meng M W, Yong L, Ze Y H. Large margin object tracking with circulant feature maps [C] // Conference on Computer Vision and Pattern Recognition, Hawaii, USA, 2017: 3065-3074
- [19] 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, Columbus, USA, 2014: 1090-1097
- [20] Danelljan M, Häger G, Khan F S, et al. Discriminative scale space tracking [J]. *Transactions on Pattern Analysis & Machine Intelligence*, 2016, 39(8): 1561-1575
- [21] 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), Las Vegas, USA, 2016: 1401-1409

Kernel correlation filtering tracking algorithm based on multi-layer deep features

Li Guoyou^{*}, Ji Zhian^{*}, Zhang Fengxu^{**}

(^{*} School of Electrical Engineering, Yanshan University, Qinhuangdao 066004)

Abstract

In view of the difficulty of dealing with the scale change, rotation and occlusion of the target in the kernel correlation filter (KCF), an anti-occlusion target tracking algorithm based on multi-layer deep features is proposed in this paper under the framework of KCF. Firstly, in frequency domain, ridge regression classifier is used to train conv3 and conv6 deep features on VGG-2048, and the confidence is obtained. The two features are weighted together to replace the original KCF's HOG features. At the same time, residual information of the first frame is introduced to obtain better position response output. Then, aiming at the occlusion problem, a response peak judgment anti-occlusion mechanism is proposed. Finally, the depth feature scale pool is established by bilinear interpolation to solve the target scale problem. The experimental results on test set (OTB-100) show that the improved algorithm can solve the target tracking problem in complex environment, and the algorithm has robustness.

Key words: kernel correlation filter (KCF) algorithm, deep feature, peak response, scale pool, robustness