

DOI:10.11830/ISSN.1000-5013.201702123



结合高斯核函数的卷积神经网络跟踪算法

郑凌云¹, 柳培忠², 汪鸿翔²

(1. 华侨大学 后勤与资产管理处, 福建 泉州 362021;
2. 华侨大学 工学院, 福建 泉州 362021)

摘要: 针对视觉跟踪中运动目标鲁棒性跟踪问题,结合高斯核函数和卷积神经网络(CNN),提出一种无需训练的卷积神经网络提取深度特征的视觉跟踪算法.首先,对初始图像进行归一化处理并聚类提取目标信息,结合跟踪过程中目标信息共同作为卷积网络结构中的各阶滤波器;其次,通过高斯核函数来提高卷积运算速度,提取目标简单抽象特征;最后,通过叠加简单层的卷积结果得到目标的深层次表达,并结合粒子滤波跟踪框架实现跟踪.结果表明:简化后的卷积网络结构能够有效地应对低分辨率、目标遮挡与形变等场景,提高复杂背景下的跟踪效率.

关键词: 视觉跟踪;卷积神经网络;高斯核函数;粒子滤波

中图分类号: TP 391 文献标志码: A 文章编号: 1000-5013(2018)05-0762-06

Convolution Neural Networks Tracking Algorithm Combined With Gaussian Kernel Function

ZHENG Lingyun¹, LIU Peizhong², WANG Hongxiang²

(1. Logistical and Asset Management Office, Huaqiao University, Quanzhou 362021, China;
2. College of Engineering, Huaqiao University, Quanzhou 362021, China)

Abstract: In view of the robustness tracking of moving targets in visual tracking, a vision tracking algorithm is proposed in this paper, which combines Gauss kernel function and convolution neural network (CNN), to extract the depth feature of the convolution neural network without training. Firstly, the initial image is normalized and the target information is extracted by clustering, and the target information in the tracking process is combined as the order filter in the convolution network structure. The Gauss kernel function is used to improve the convolution operation speed, extract the simple abstract feature of the target, and then superimpose the convolution results of the simple layer to get the depth of the target. Finally, we combine particle filter tracking framework to achieve tracking. The results show that the simplified convolution network structure can effectively cope with low resolution, target occlusion and deformation and so on, and improve the tracking efficiency in complex background.

Keywords: visual tracking; convolutional neural network; Gauss kernel function; particle filter

收稿日期: 2017-03-28
通信作者: 柳培忠(1976-),男,副教授,博士,主要从事仿生图像处理技术,智能算法的研究. E-mail: pzliu@hqu.edu.cn.
基金项目: 国家自然科学基金资助项目(61203242);福建省物联网云计算平台建设基金资助项目(2013H2002);华侨大学研究生科研创新能力培育计划资助项目(1511422004).

视觉跟踪是机器视觉领域的研究热点,目前已经取得很多研究成果,但在实际应对各类复杂场景时仍面临很大挑战,如何实现更加鲁棒和准确的跟踪仍然是研究的核心^[1]. 视觉跟踪算法主要分为判别式和生成式两类^[2]. 判别模型将跟踪问题建模为一个二分类问题,用以区分前景和背景. Babenko 等^[3]提出多示例学习算法,针对训练样本不足,引入多示例学习机制,有效抑制跟踪过程中的漂移问题;钟必能等^[4]结合选择性搜索和粒子滤波提取候选粒子,融合卷积深信度网络和卷积神经网络,有效地应对遮挡、光照等复杂场景. 生成模型不考虑背景信息直接为目标进行建模. Mei 等^[5]提出 L1 跟踪系统,用稀疏方法表示跟踪目标,但算法运算复杂度高;Zhang 等^[6]提出压缩跟踪,用一个稀疏的测量矩阵提取特征,建立一个稀疏健壮的目标表现模型,提高算法的实时性与鲁棒性. 近几年,相关滤波跟踪算法在目标领域快速发展. David 等^[7]将相关滤波引入跟踪领域,设计一个误差最小平方和滤波器(MOSSE),在跟踪过程中寻找目标最大响应值来实现跟踪. Henriques 等^[8]通过循环偏移构建分类器的训练样本,引入方向梯度直方图(HOG)和颜色、灰度多通道特征,提高了算法的速度和准确性. 传统跟踪算法大多使用视频序列浅层的像素级特征,无法很好应对复杂场景. 本文结合卷积神经网络强大的特征提取功能,设计一种无需训练的卷积神经网络特征提取方法,利用核函数加速卷积运算,结合粒子滤波框架下实现一种快速卷积神经网络跟踪算法.

1 相关工作

1.1 CNN 特征提取结构

卷积神经网络(CNN)的网络结构类似于生物神经网络,采用权值共享、时空下采样和局部感受野的思想降低了网络复杂度,使得 CNN 在处理高维图像时更具优势. 卷积神经网络具有多层性,每层由多个二维平面特征映射层组成,每个映射层由多个独立的神经元组成,如图 1 所示. 提取卷积特征如下;首先,对原始输入图像进行取片操作获取大量小的局部块;然后,应用卷积网络模型对局部块进行训练,得到结果为各个卷积层的神经元卷积滤波器;最后,与新输入的样本图像进行卷积滤波,提取样本图的抽象卷积特征,从而得到原始图像的深度特征^[9].

1.2 基于卷积神经网络的跟踪算法

基于卷积神经网络的跟踪算法,通过大量的前期预训练来训练神经网络提取深度特征,面临样本缺少、算法时间复杂度高、硬件运行平台要求高,以及跟踪实时性低等显著缺点. 针对以上问题, Wang 等^[10]提出 DLT(deep learning tracking)算法首次将深度学习算法应用到跟踪领域,在自然图像数据集上使用栈式降噪自编码器离线预训练来获得通用的物体表征能力,利用粒子滤波框架并在跟踪过程中更新自编码器实现跟踪. Zhang 等^[11]结合人脑视觉处理系统,简化了卷积网络结构,使用目标区域中随机提取的归一化图像块作为卷积神经网络的滤波器,从而实现了不用训练卷积神经网络的快速特征提取. Nam 等^[12]提出训练方法和训练数据交叉运用的思路,在检测任务重点关注背景中的难点样本显著减轻了跟踪器漂移问题.

在特征提取时,文中结合卷积神经网络,能够获得一定的平移、尺度和旋转不变性,结合文献[11]提出的无需训练的卷积特征提取方法,采用两层前馈处理方式简化卷积网络结构;然后,通过分层滤波器卷积抽取目标的高维抽象特征,利用高斯核函数进行运算加速,提出一种快速卷积神经网络跟踪算法.

2 结合核函数的卷积网络跟踪算法

2.1 核函数卷积

核函数可以将低维空间中不能线性分割的点集,通过转化为高维空间中的点集,变为线性可分的. 由于高斯核函数的可分离性,二维高斯函数卷积可以分两步进行. 首先,将图像与一维高斯函数进行卷积;然后,将卷积结果与方向垂直的相同一维高斯函数卷积. 文献[8]采用子窗口高斯核函数 $k(x,$

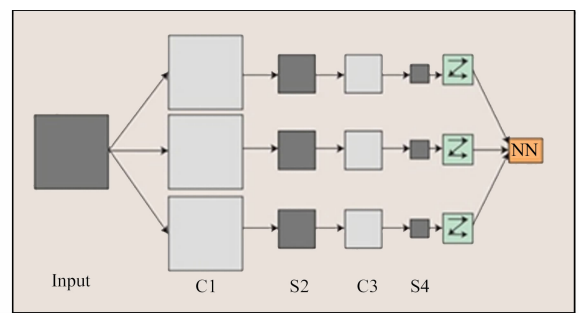


图 1 卷积特征提取结构
Fig. 1 Convolution feature extraction structur

x''))的快速计算方法,而文中的卷积运算采用高斯核函数进行变换加速运算.

设核函数 $k(x, x'')$, 则权重可表示为 $v = \sum_i \alpha \phi(x_i)$, 系数向量为 α , 其元素为 α_i , 可以使用线性函数的优化方法求解 α . 因此, 核正则化最小二乘分类器具有一个简单的闭式解, 即 $\alpha = (K + \lambda I)^{-1} y$. 其中: K 是核函数矩阵, 矩阵元素为 $K_{i,j} = k(x_i, x_j)$; I 是单位矩阵; 向量 y 的元素为 y_i . 由于 K 是循环矩阵, 则转换到离散傅里叶变换(DFT)域. 即

$$\hat{\alpha}^* = \hat{y} \times (\hat{k}^{x,x} + \lambda)^{-1}.$$
(1)

式(1)中: $\hat{k}^{x,x}$ 是核函数矩阵 K 的第一行元素组成的向量, 符号“ \sim ”表示傅里叶变换. 由此可知, 核正则化最小二乘分类器闭式解式可利用快速傅里叶变换(FFT)得到.

2.2 特征提取

利用卷积网络设计一个分层的目标表示结构. 在第 1 帧中, 将目标归一化到 $n \times n$ 大小, 从中随机选取一部分重叠的小图像块组成一系列滤波器; 然后, 将用于同第 t 帧候选目标框中所有抽取的图像块进行卷积; 最后, 可得到一系列局部特征的映射, 从而提高候选目标全局特性的鲁棒性.

1) 简单层特征. 通过预处理将图像归一化到 $n \times n$ 大小, 对目标区域利用 $w \times w$ 大小的滑动窗口进行采样, 得到长度为 L 的图像块组 X ; 用 k -means 聚类方法从 $L = (n - w + 1) \times (n - w + 1)$ 个图片块中聚类出 d 个图像块滤波器作为卷积核, 记 $F = \{F_1, F_2, \dots, F_d\} \subset X$. 输入图像 I 对应的响应式为

$$S_i^o = F_i^o \otimes I, \quad S_i^o \in \mathbf{R}^{(n-w+1) \times (n-w+1)}.$$
(2)

2) 复杂层特征. 为了加强对目标的特征表达, 将 d 个简单层的特征进行堆叠, 构成一个三维张量来表示目标的复杂层特征, 记作 $C \in \mathbf{R}^{(n-w+1) \times (n-w+1) \times d}$. 这种特征具有平移不变特性. 由于图像归一化后, 使得特征对目标的尺度具有鲁棒性, 且复杂层特征保留不同尺度目标的局部几何信息.

2.3 粒子滤波

基于粒子滤波框架, 设第 t 帧时总的观测序列为 $O_t = \{o_1, \dots, o_t\}$, 则跟踪主要是求出后验概率 p 的最大值. 根据贝叶斯理论, 可得

$$p(S_t | O_t) \propto p(O_t | S_t) \int p(S_t | S_{t-1}) p(S_{t-1} | O_{t-1}) dS_{t-1}.$$
(3)

式(3)中: $S_t = [x_t, y_t, s_t]^T$, x_t, y_t 为目标的位置, s_t 为尺度参数; $p(S_t | S_{t-1})$ 为运动模型, 用于根据第 $t-1$ 帧的位置预测第 t 帧的位置; $p(S_t | O_t)$ 为观测模型, 用于评估观测结果 O_t 与目标的相似性.

观测模型通过测量样本通目标之间的相似度计算, 有 $p(O_t | s_t^i) \propto \exp(-|\text{vec}(C_t) - \text{vec}(C_t^i)|_2^1)$. 于是, 整个跟踪过程为

$$\hat{S}_t = \operatorname{argmax}_{\{s_t^i\}_{i=1}^N} p(O_t | s_t^i) p(s_t^i | \hat{s}_{t-1}).$$
(4)

2.4 跟踪算法

基于卷积特征, 结合粒子滤波, 提出核函数卷积神经网络跟踪算法.

1) $T=1$, 第一帧.

- 步骤 1 对第一帧图片序列进行初始化与预处理操作.
- 步骤 2 标记出需要跟踪的目标.
- 步骤 3 提取初始状态滤波器.
- 步骤 4 利用式(1), (2)核函数卷积学习到目标的抽象特征.
- 步骤 5 构建目标的特征模板.
- 步骤 6 通过粒子滤波获取候选粒子.

2) For $T=2$, 到最后一帧.

- 步骤 1 对输入图片进行初始化与预处理操作.
- 步骤 2 利用式(1), (2)核函数卷积, 学习到目标的抽象特征.
- 步骤 3 通过动态模型对每个粒子进行预测.
- 步骤 4 通过观测模型计算粒子权值.
- 步骤 5 根据式(4)权值最大的粒子作为目标物体.
- 步骤 6 通过相应的权值对粒子重采样.

End for.

3 实验结果与分析

采用定性分析与定量分析相结合,对算法进行分析验证. 利用 MATLAB2014a 编程环境,PC 配置为 Inter Core i3-3220,3.3 GHz,8 G 内存,在测试集 Database OTB2013 中进行仿真分析^[2,13-16]. 仿真参数设置如下:滤波网络尺寸 6×6,滤波器数 100,归一化尺寸 32×32,学习因子设置为 0.95. 粒子滤波器的目标状态的标准偏差设置如下: $\sigma_x=4,\sigma_y=4,\sigma_s=0.01$,使用 $N=300$ 个粒子.

3.1 定性分析

限于论文篇幅,仅给出几组代表性的跟踪实验结果,如图 2 所示. 图 2 中:红色实线为文中所提算法;绿色实线为压缩跟踪(compressive tracking,CT)算法^[6];蓝色实线为核化相关滤波器(kernelized correlation filter,KCF)算法^[8];黑色实线为基于卷积网络的跟踪(CNT)算法^[11].

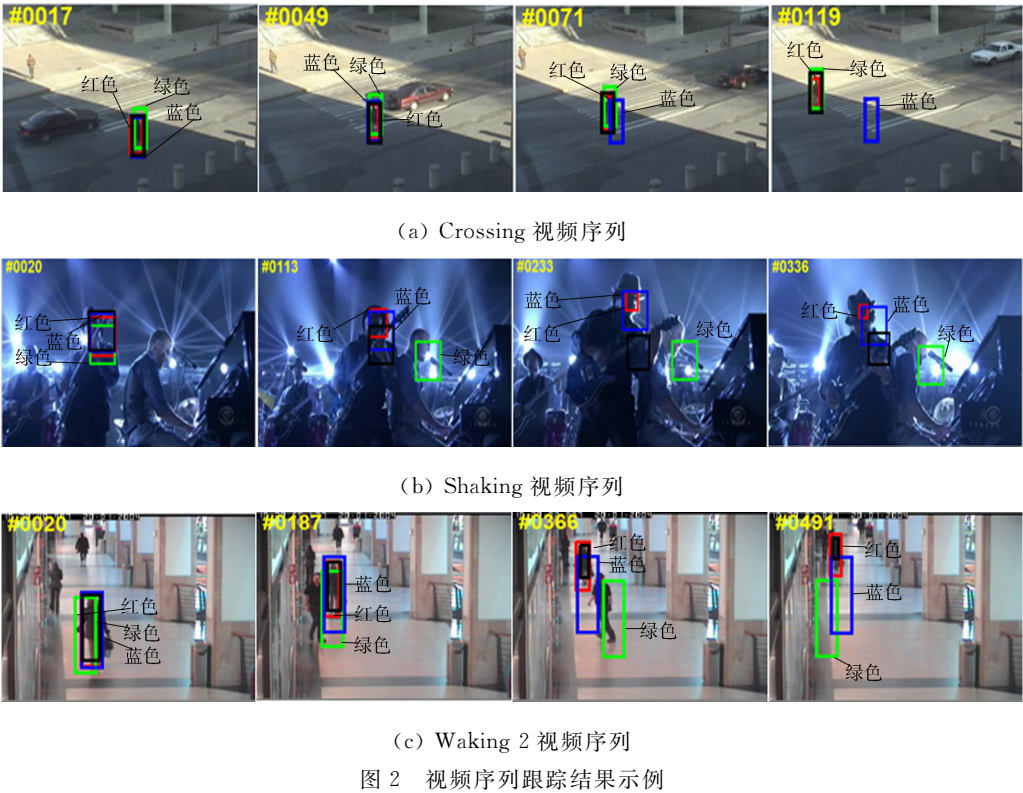


Fig. 2 Examples of the tracking results on video sequences

由图 2 可知:第一组序列 Crossing 在背景干扰并伴随着光照变化,目标过马路途中一直存在目标形变与相应尺度变化,文中所提算法在所有算法中取得了最准确的跟踪效果;第二组序列 Shaking 目标始终伴随强烈的光照与尺度变化,跟踪过程 CT 算法与 CNT 算法效果不佳,文中所提算法与 KCF 相对比较准确;第三组序列 Waking 2 是在低分率场景下的监控视频,目标在行进过程中会发生一定的遮挡和相应尺度变化.此时,采用卷积神经网络进行特征提取,能都提取到深层次更加鲁棒的特征,所以文中所提算法在遮挡后仍能实现有效跟踪.

3.2 定量分析

为了测试算法性能,采用如下几个评价准则. 1) 中心位置误差(CLE),即目标的中心位置与标准中心位置的欧式距离的误差, $CLE = (\sum_{i=1}^n C_i - C_i^r)/n$. 其中: n 是图片序列的帧数; C_i 为目标中心位置; C_i^r 为标准中心位置. 2) 距离精度(DP),即中心误差小于一个给定阈值的帧的相对数量, $DP=m/n$, m 为 CLE 值小于某个固定阈值的图片序列帧数, n 为总帧数,当算法跟踪的中心位置与目标真实中心位置的误差不大于阈值时,就认为跟踪准确,并给出最终跟踪准确占比.

4 种算法在各个测试序列中的平均中心位置误差和距离精度(此处取阈值为 20 px),如表 1 所示.

表 1 中:数值加粗且加下划线的代表最优性能,数值加粗表次优性能.由表 1 可知:文中所提算法在位置中心误差与距离精度两项评判指标中均表现出最优效果.

表 1 4 种算法在各个测试序列中的评价指标

Tab. 1 Evaluation index of four algorithms in each test sequence

视频序列	文中算法		CT 算法		KCF 算法		CNT 算法	
	CLE	DP	CLE	DP	CLE	DP	CLE	DP
Crossing	<u>1.640</u>	1.000	5.230	1.000	2.520	1.000	2.790	1.000
Dancer 2	<u>6.110</u>	1.000	9.510	0.973	6.410	1.000	7.140	1.000
David 2	<u>2.040</u>	1.000	15.400	0.780	2.360	1.000	2.640	1.000
Shaking	<u>16.700</u>	<u>0.712</u>	98.000	0.011	105.000	0.025	39.800	0.011
Walking 2	<u>3.230</u>	1.000	58.500	0.396	28.600	0.440	3.860	1.000

4 种算法的具体每一帧的中心位置误差(ϵ)曲线图,如图 3 所示.图 3 中: N 表示帧号;纵坐标表示算法跟踪中心位置与真实目标中心位置的偏离误差.由图 3 可知:在跟踪过程中,当目标发生遮挡与形变导致目标丢失后,会出现 CLE 急剧升高的变化.文中所提的跟踪算法具有较好的准确性与鲁棒性,误差曲线在测试序列中呈现较低且相对平滑的趋势.

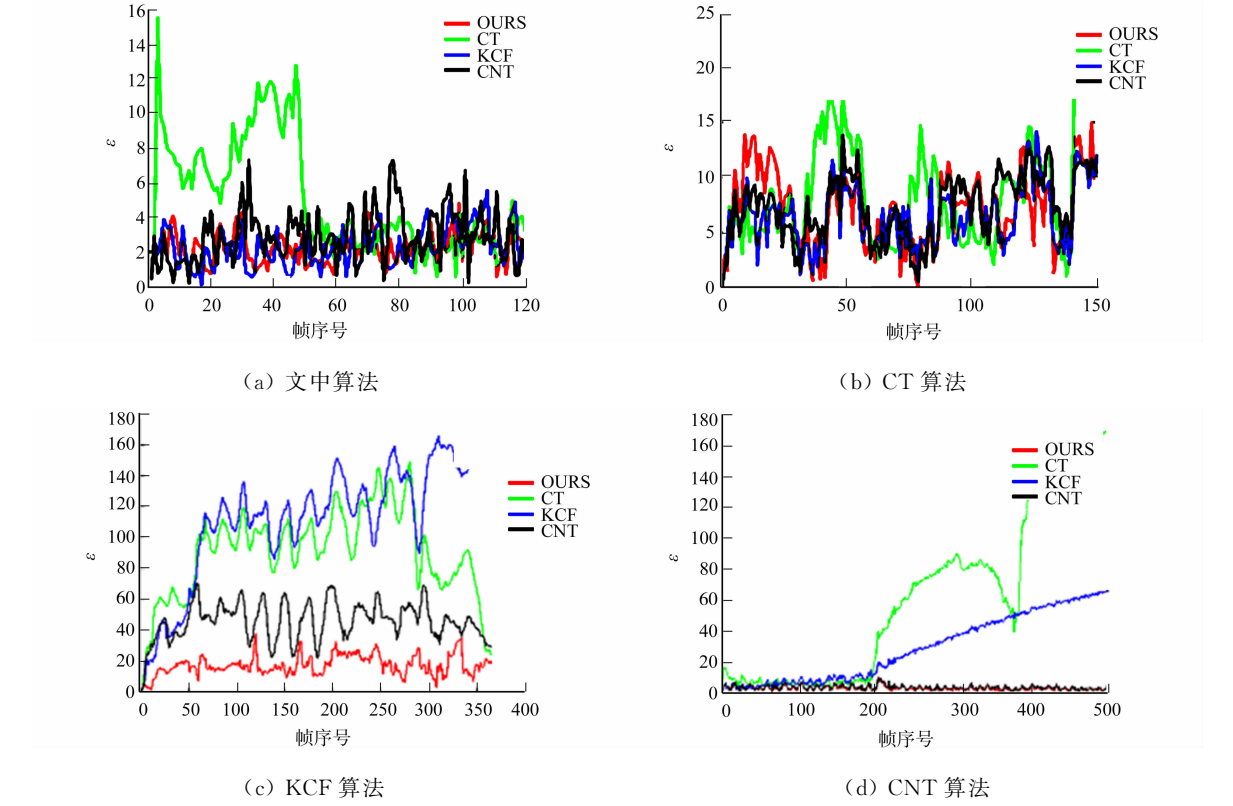


图 3 4 种算法的中心位置误差曲线图

Fig. 3 Central position error graphs for four algorithms

4 结束语

采用简化后的卷积网络结构提取目标的深度抽象特征,并引入高斯核函数加速计算,结合粒子滤波框架,在目标形变、遮挡、低分辨率等场景下,取得较好跟踪效果.因为文中的特征提取方式采用卷积神经网络特征,加速后仍存在算法时间复杂度高等问题,所以,所提算法在快速运动、目标出界等场景下仍面临很大挑战,在今后的工作中,将主要致力于解决此类场景的跟踪问题.

参考文献:

[1] 杨戈,刘宏.视觉跟踪算法综述[J].智能系统学报,2010,5(2):95-105. DOI:10.3969/j.issn.1673-4785.2010.02.

- 001.
- [2] WU Yi,LIM J,YANG M H. Online object tracking: A benchmark[C]//IEEE Conference on Computer Vision and Pattern Recognition. Portland;IEEE Press,2013;1354-1362. DOI:10. 1109/CVPR. 2013. 312.
- [3] BABENKO B,YANG M H,BELONGIE S. Robust object tracking with online multiple instance learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2011,33(8):1619-1632. DOI:10. 1109/TPAMI. 2010. 226.
- [4] 钟必能,潘胜男. 选择性搜索和多深度学习模型融合的目标跟踪[J]. 华侨大学学报(自然科学版),2016,37(2):207-212. DOI:10. 11830/ISSN. 1000-5013. 2016. 02. 0207.
- [5] MEI Xue,LING Haibin. Robust visual tracking using L1 minimization[C]//IEEE International Conference on Computer Vision. [S. l.]:IEEE Computer Society,2009;1436-1443. DOI:10. 1109/ICCV. 2009. 5459292.
- [6] ZHANG Kaihua,ZHANG Lei,YANG Minghsuan. Real-time compressive tracking[C]//Proceedings of the 12th European Conference on Computer Vision- Volume (III). Berlin;Springer-Verlag,2012,7574(1):864-877. DOI:10. 1007/978-3-642-33712-3_62.
- [7] DAVID S,J ROSS,BRUCE A. Visual object tracking using adaptive correlation filters[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San Francisco;IEEE Press,2010;2544-2550. DOI:10. 1109/CVPR. 2010. 5539960.
- [8] 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. DOI:10. 1109/TPAMI. 2014. 2345390.
- [9] 余凯,贾磊,陈雨强,等. 深度学习的昨天、今天和明天[J]. 计算机研究与发展,2013,50(9):1799-1804.
- [10] WANG Naiyan,YEUNG D Y. Learning a deep compact image representation for visual tracking[C]//27th Annual Conference on Neural Information Processing Systems. Nevada;Curran Associates Inc,2013;809-817.
- [11] ZHANG Kaihua,LIU Qingshan,WU Yi,*et al.* Robust tracking via convolutional networks without learning[J]. IEEE Transactions on Image Processing,2015,25(4):1779-1792.
- [12] NAM H,HAN B. Learning multi-domain convolutional neural networks for visual tracking[C]//IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas;IEEE Press,2016;4293-4302. DOI:10. 1109/CVPR. 2016. 465.
- [13] COMAMICIU D,RAMESH V,MEER P. Real-time tracking of non-rigid objects using mean-shift[C]//Proceedings of the IEEE Conference Computer Vision and Patter Recognition. Hilton Head Island;IEEE Press,2000;142-149. DOI:10. 1109/CVPR. 2000. 854761.
- [14] ISARD M,BLAKE A. Condensation-conditional density propagation for visual tracking[J]. International Journal of Computer Vision,1998,29(1):5-28. DOI:10. 1023/A:1008078328650.
- [15] ROSS D A,LIM J,LIN R S,*et al.* Incremental learning for robust visual tracking[J]. International Journal of Computer Vision,2008,77(1):125-141. DOI:10. 1007/s11263-007-0075-7.
- [16] WU Y,LIM J,YANG M H. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2013,9(4):2411-2418. DOI:10. 1109/TPAMI. 2014. 2388226.

(责任编辑: 陈志贤 英文审校: 吴逢铁)