

基于显著性特征的选择性目标跟踪算法

丁新尧, 张 鑫

(华南理工大学电子与信息学院, 广东广州 510640)

摘 要: 针对长期目标跟踪算法中目标部分遮挡甚至消失情况下的目标有效跟踪问题, 提出了一种融合了目标显著性特征的选择性跟踪算法. 首先, 为了有效抑制背景信息的干扰, 综合 HOG 特征以及颜色统计特征的特点提出了前景概率图来实现增强目标显著性抑制背景干扰的效果. 其次, 为了减少跟踪漂移和解决重度照明和遮挡等挑战性场景中的跟踪失败问题, 引入了具有筛选条件的选择性跟踪和检测框架, 用以控制检测器的激活以及最终结果的选择. OTB2013 数据集上的实验结果证明, 本文算法可以取得 91.1% 的总体准确率以及 67% 的总体成功率, 结果优于大部分跟踪算法.

关键词: 前景概率图; 条件检测机制; 跟踪置信度; 特征显著性

中图分类号: TP391.41 **文献标识码:** A **文章编号:** 0372-2112 (2020)01-0118-06

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.01.014

Visual Tracking with Salient Features and Selective Mechanism

DING Xin-yao, ZHANG Xin

(School of Electronics and Information Engineering, South China University of Technology, Guangzhou, Guangdong 510640, China)

Abstract: In the long time tracking, object representation and occlusion handling are two important challenges. We propose a selective tracking and detection framework in which a new probabilistic object-enhanced feature is integrated. Firstly, we propose a foreground probability map to enhance the target and weaken the surrounding background. Secondly, we introduce the selective tracking and detection framework that has two sets of conditions to control the detector activation and final result selection. We have evaluated our methods on the popular benchmark OTB2013 dataset. The algorithm achieves an overall accuracy of 91.1% and a success rate of 67%, which demonstrates that our algorithm performs favorably compared with other state-of-the-art methods.

Key words: foreground probability map; condition detection mechanism; tracking confidence; salient features

1 引言

目标跟踪是计算机视觉的基本问题, 其主要目的是在连续的图像序列之间, 基于目标位置、速度、形状、纹理、色彩等特征对兴趣目标进行匹配, 在图像序列中完成目标定位, 从而获得目标的运动轨迹和形变参数^[1]. 目标跟踪在民用和军事的许多领域^[2-5]中都具有极为广泛的应用前景.

在连续视频帧的目标跟踪中, 当目标出现光照变化和遮挡等情况时, 目标的表观特征会发生变化进而导致跟踪漂移甚至目标丢失问题. 针对这些挑战, 本文提出一种基于显著性特征的目标跟踪算法 (selective target tracking algorithm based on salient features,

STSF). 对于目标周围的背景信息以及目标相似性背景信息可能导致跟踪失败这一点, 利用提出的前景概率图 (Foreground Probability Map, FP Map) 来增强前景和背景的对比性. 具体地, 用前景概率图和 HOG 特征进行点乘来获得较为鲁棒的目标表观特征. 为了解决跟踪漂移以及目标遮挡问题, 提出了一个条件检测机制来控制检测器的启动以及跟踪结果的选取. 条件检测机制包括两个决策模块, 分别为检测决策模块和结果选择决策模块. 其中检测决策模块用来决定什么情况下启动检测器, 结果选择决策模块是用于决定启动了检测器之后如何进行最后的结果选择. 这两个模块的配合使得算法更适合于长期跟踪.

本文的主要贡献: (1) 在目标特征增强方面, 提出

一种前景概率图来增强前景特征抑制背景特征,突出目标特征的显著性;(2)为了解决跟踪漂移和部分目标遮挡问题,提出一种条件检测机制来辅助跟踪器进行鲁棒性跟踪;(3)通过在 OTB2013^[6]数据库上的实验证明了算法的鲁棒性,此外,算法的跟踪速度是相对比较快的。

2 相关工作

2.1 长期目标跟踪算法

为了解决跟踪的过程中目标遮挡甚至短暂性的消失又重现的问题,一系列长期目标跟踪算法相继提出。MUSTer 算法^[7]基于人脑记忆的启发,提出了一个长期记忆机制和短期记忆机制相互配合的方法来得到一个比较鲁棒的跟踪结果。MEEM^[8]算法提出在跟踪模型进行更新的时候专门学习了一个专家评测模型,在实际的过程中利用专家模型对跟踪结果进行分数评判,进而防止目标漂移以及跟踪丢失问题。

2.2 SiamFC 网络

深度学习在目标跟踪中的应用^[9,10]也是比较广泛的。全卷积孪生网络(Siamese Network)^[11]的主体架构

是由两个全卷积网络通道组成,通过求取两个特征通道的相关性确定最终的目标位置。其中,全卷积网络仅作为基准图像和候选图像的特征提取器,经过网络以及相关性求取便可以得到最后的置信图。再根据置信图的最大值的坐标反推获取跟踪目标的位置信息。

3 基于显著性特征的选择性跟踪算法

STSF 算法分为三个部分:基础跟踪框架、前景概率图作用下的特征增强以及条件检测机制。

3.1 基础跟踪框架

基础的跟踪框架包括跟踪器、检测器、检测条件和判决条件四个模块。在跟踪器模块中,利用前景概率图在 BACF^[12]这一基础跟踪框架上提取目标的增强型特征,突出前景特征抑制背景干扰。如图 1,在条件检测这一模块中,跟踪置信度和前景概率图的置信度都高于标准红色虚线的时候,代表结果跟踪可信;否则,检测器将会启动。对于检测器,利用全卷积孪生网络的特性来进行目标的重检测。最后对于检测器和跟踪器的结果,判决条件会对两者进行特征提取获取其置信度,然后进行最终结果的选择。

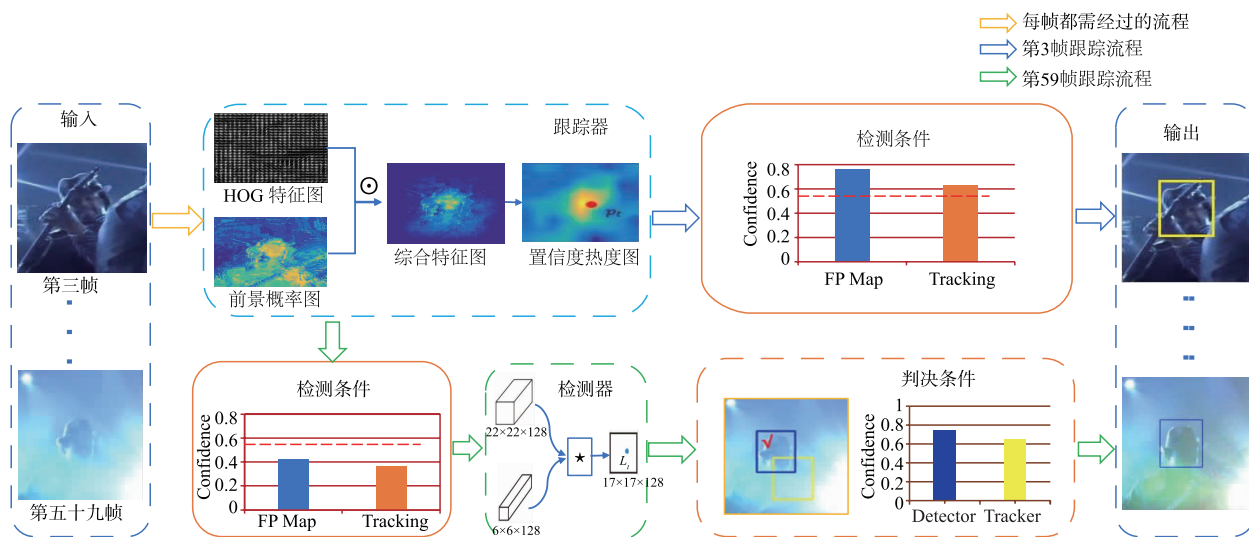


图1 STSF算法的整体流程框架

3.2 前景概率图下的目标特征增强

由于跟踪目标的周围存在大量的背景干扰信息,特别是和目标具有相似特征的物体时,算法很容易出现跟踪漂移。本文算法利用颜色统计模型得到的颜色前景概率图,凸显目标特征减少模型对背景干扰信息的关注度,进而尝试解决部分遮挡和背景干扰等问题来辅助目标跟踪。

前景概率图是基于图片颜色信息求取的,给定上一帧的目标区域 Z , 区域中的每个像素都有红绿蓝三种颜色通道,将其像素值变化范围统一归一化到 $[1, 32]$ 。

这样以来每个像素点的三个颜色通道一起就会有 $32 \times 32 \times 32$ 种可能性。假定在 $t-1$ 帧目标的区域大小是 $w \times h$, 可以将目标周围 $(w-k) \times (h-k)$ 所框定的区域用 Z_f 表示,代表前景区域。在每帧中划定的检测区域与 $(w+k) \times (h+k)$ 之间的区域用 Z_b 表示,代表的是背景区域。其中 k 等于 $\alpha * (w+h)/2$, α 是一个用来调整得到合适的前景和背景区域的参数。为了计算前景概率图,需要从前一帧 $t-1$ 中分别统计出前景和背景的两个字典模型 $\beta_{t-1}(F)$ 和 $\beta_{t-1}(B)$ 。字典模型可以按如下形式计算:

$$\beta_{i-1}^j(F) = \frac{\sum_{i=1}^{N_f} \varphi(z_i)}{N_f}, z_i \in Z_f \quad (1)$$

$$\beta_{i-1}^j(B) = \frac{\sum_{i=1}^{N_b} \varphi(z_i)}{N_b}, z_i \in Z_b \quad (2)$$

其中 z_i 代表的是每个像素点三个通道的像素值, j 是一个三维序列向量. N_f 和 N_b 分别表示的是前景和背景区域的像素个数. 函数 φ 的作用就是判断像素向量 z_i 和序列向量 j 是否是一样的, 如果一样就会返回 1, 否则返回 0. 然后就可以得到两个 $32 \times 32 \times 32$ 维的前景和背景的字典, 字典每一帧都需要根据一定的更新率进行动态更新.

给定当前帧 I_t , 依据前一帧的跟踪结果可以得到输入检测图像块 Z_t 大小为 $(w+k) \times (h+k)$. 然后依据第 $t-1$ 帧中前景和背景两个字典可以计算出两个概率模型 $\omega(B)$ 和 $\omega(F)$, 其中 $\omega(F)$ 代表的是根据 $\beta_{t-1}(F)$ 统计出的在第 t 帧中检测区域中的各个像素位于前景区域的概率图, 对应地 $\omega(B)$ 代表的是在第 t 帧中检测区域中的各个像素是属于背景区域的像素的概率图. 所以对应的前景概率图可以表述为:

$$f_{\text{FPM}} = \frac{\omega(F)}{\omega(F) + \omega(B)} \quad (3)$$

其中 $\frac{\cdot}{\cdot}$ 代表的是对应元素相除, $\cdot + \cdot$ 代表的是对应元素相加. 在颜色特征基础上, 前景概率图对于前景和背景的概率分布有一个准确的描述. HOG 特征可以对边缘特征有一个清晰的描述, 因此算法在前景概率图和 HOG 图之间进行了点乘, 如下:

$$f_{\text{FINAL}} = f_{\text{FPM}} \odot f_{\text{HOG}} \quad (4)$$

其中 f_{FPM} 代表前景概率图, f_{HOG} 代表的是 HOG 特征图, \odot 代表对应元素点乘.

3.3 条件检测机制

对于长期目标跟踪中目标丢失和再现的问题, 引入检测器主要是为了在目标遮挡和消失的情况下扩大算法的感受野, 保证目标大概率在检测范围内. 当目标再次出现时, 结合增强型特征突出目标特征, 辅助检测器快速检测到目标, 进而重新跟踪. 由于前景概率图对于模糊变形等情况的鲁棒性比较好, 比较适合用来反映跟踪状态. 根据前景概率图得到的置信值结合式(3)可以表述为:

$$S_{\text{FPM}} = \max f_{\text{FPM}} \quad (5)$$

如果 $S_{\text{FPM}} < T_c S_{\text{FPM}}^0$, 那么表示目标可能已经跟丢了或者跟踪场景经历了比较大的变化. T_c 代表的是颜色阈值, S_{FPM}^0 表示的是第一帧计算出来的前景概率图置信度. 除了前景概率置信度外, 还引入了另一个跟踪置信度 S_T 作为一个补充. 相比较来说, 跟踪置信度对跟踪的效果反映更加直接一些. 在第 t 帧的跟踪置信度可以表述为以下形式:

$$R_t = F^{-1}(\hat{g}_{t-1} \odot \hat{f}_{\text{FINAL}}) \quad (6)$$

$$S_T = R_t(x_T, y_T) \quad (7)$$

其中 \hat{g}_{t-1} 是一个 BACF^[7] 中的频域中的相关滤波模型, 经过 ADMM 算法迭代得到的, 详细推导可见文献[7]. \hat{f}_{FINAL} 是在第 t 帧的最终频域特征. R_t 是最终的跟踪置信图, 而 (x_T, y_T) 代表的是由图 1 中跟踪器模块得到的目标的中心位置. S_T 就对应的是目标位置处的置信度值. 对于这个置信度来说, 当 $S_T < T_t S_T^0$ 时, 检测器将被激活, T_t 代表的是跟踪阈值, S_T^0 代表第一帧的跟踪置信度. 当目标旋转或者在检测区域之外的時候, 跟踪置信度可能会下降, 然后就激活检测器. 跟踪置信度和前景概率置信度是决定是否启动检测器的条件, 检测器在具有挑战性的场景下会被激活并提供一个检测结果.

$$M_t = F^{-1}(\hat{g}_{t-m} \odot \hat{f}_{\text{FINAL}}) \quad (8)$$

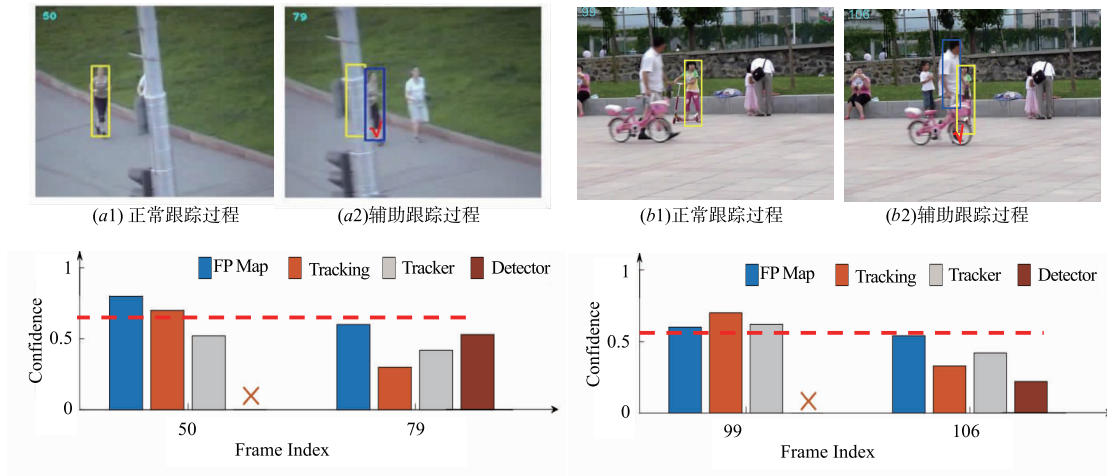


图2 条件检测机制实验效果对比图

$$(x_F, y_F) = \begin{cases} (x_D, y_D), & M_i(x_D, y_D) > M_i(x_T, y_T) \\ (x_T, y_T), & \text{otherwise} \end{cases} \quad (9)$$

其中 M_i 和式(6)中 R_i 的区别在于更新频率, R_i 是每一帧都进行更新的. 然而, 仅仅在满足 $M_i(x, y) > T_b$ 时, 算法才对 T_b 进行更新, 其中 T_b 是一个模型阈值. 变量 m 表示这里有连续 m 帧不满足更新条件, 那 \hat{g} 就不更新一直保持 m 帧前的值不变. 相比较 R_i 而言, 由于这种方法模型更新的不够及时, 不能实时捕捉到目标的表现变化, 得到的 M_i 不适合短期跟踪. 但是, 这样更新带来的好处就是模型很少引入干扰和污染, 对目标的描述比较准确, 所以适合长期跟踪中的检测部分. 此外, 在交叉遮挡性数据集进行了方案有效性验证, 如图 2 所示, 检测器可以很好地辅助以及纠正跟踪器的跟踪结果.

4 实验结果与分析

为了证明 STSF 算法的各个部分以及算法整体的有效性, 本文算法与其他九个效果很好的算法在 OTB2013 视频数据库上做了跟踪效果的有效性对比. 实验是在 Intel(R) Core(TM) i7-4790 CPU 3.6GHZ 处理器, 内存为 32G 的 64 位操作系统的计算机上, 使用 MATLAB R2016a 进行的. 算法对于 OTB2015 的 100 个视频帧的平均跟踪帧率为 18.4FPS.

4.1 整体算法跟踪性能对比实验

本文算法在 OTB benchmark 上与九种跟踪算法在 OTB2013 数据集上进行实验测试. 实验比较的九种算法分别为 SiamFC^[11], BACF^[12], KCF^[13], Staple^[14], LCT^[15], MEEM, SAMF^[16], DeepSRDCF^[17], SRDCF^[18]. 其中 DeepSRDCF^[17], SRDCF^[18] 以及 SiamFC 利用的是深度学习的方法, MEEM 以及 LCT 是长期目标跟踪算法, 其余的利用的是相关滤波方法.

在 OTB2013 数据库中, 利用的是常规测试 (OPE) 的方法对十种跟踪算法结果进行了比较, 如图 3. 准确率的标准是与标签中目标区域相比像素精度差在 20 个像素以内的图片帧数与总帧数之比, 即图 3 中左图横轴阈值 20 处对应的准确率. 成功率的值是指右图中各算法对应曲线与横轴和纵轴围成的面积. 相对比与基础模型框架 BACF, 算法模型有了明显的效果提升, 在十种算法中达到了最好的效果.

4.2 算法鲁棒性实验与分析

为了进一步测试算法在各种场景下的鲁棒性, 在图 4 中, 利用 TB50^[12] 数据库中背景干扰、运动模糊和快速运动三个数据属性的视频对 6 种算法进行分属性实验测试对比. 下面对抽出的三个视频帧进行结果分析.

在图 4 中, 自上而下的三个视频数据分别为 Football, Jump, Skiing.

Football: 主要用于测试背景出现相似物干扰时的算法的跟踪效果, 如图 4 所示. 当目标与背景中的相似物越来越远共同进入目标检测视野时, 大部分算法的辨识度不够, 很容易混淆跟错背景目标. 在前景概率图的作用下, 目标的特征显著性更加突出, 使得算法对于目标跟踪更加准确.

Jump: 主要用于进行背景模糊状况下的算法跟踪效果的测试. 该视频帧中由于目标的快速运动, 使得目标一直处于模糊状态, 分辨率较低. 此时大部分算法容易发生跟踪漂移, 而本文算法以及 KCF 可以很好的跟踪目标.

Skiing: 此视频中是一个高速运动的滑雪运动员在高速运动滑翔. 这种情况下, 由于目标自身的快速运动, 使得相邻帧间的目标位置变化较大, 算法很容易发生跟踪丢失. 由图 4 可以看出本文算法依然可以进行很好的跟踪.

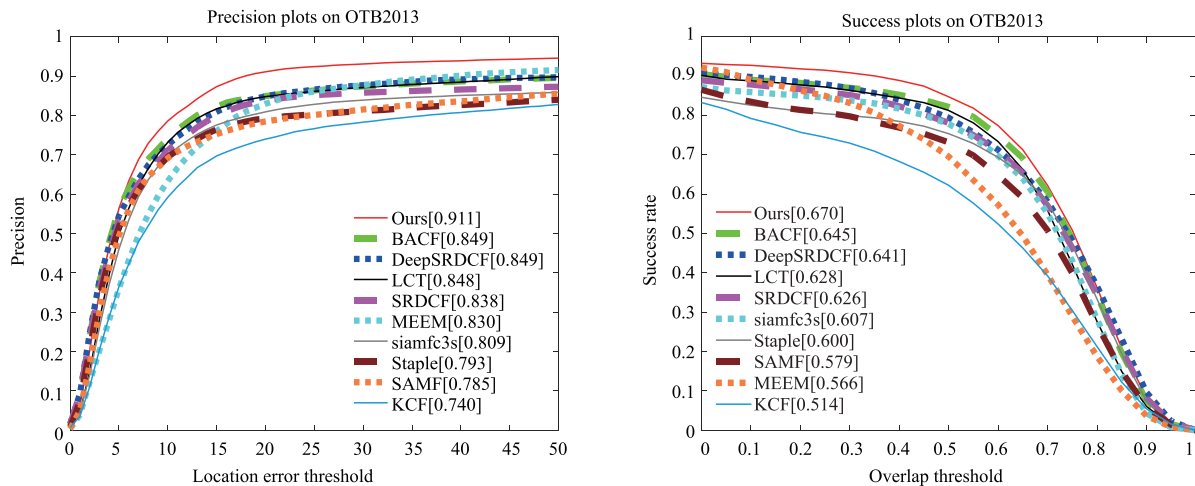


图3 OTB2013的结果对比图

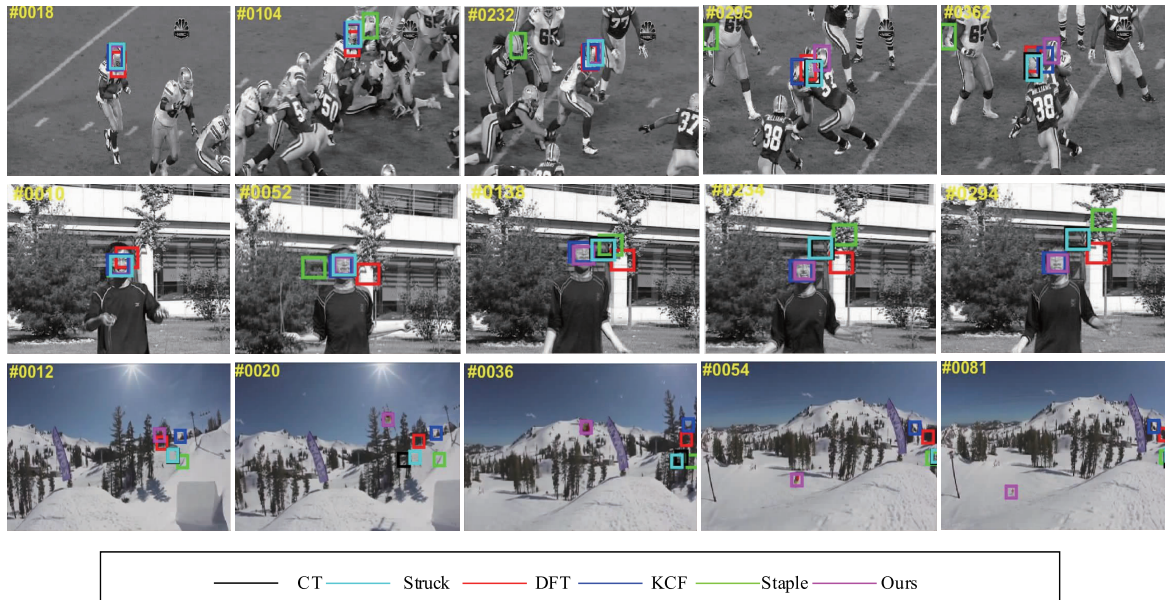


图4 部分视频序列跟踪结果

4.3 对比试验分析

通过以上数据集的实验对比,充分证明了 STSF 算法的整体有效性. 其中 DeepSRDCF^[17] 以及 SRDCF^[18] 等深度学习方法,虽然对于目标的特征提取比较充分,但是缺少抑制背景增强前景的这种对比增强型特征. 因此,在抗干扰能力方面本文算法更加鲁棒. 另一方面,立足于长期目标跟踪,对比于 MEEM 算法的多专家模型,本文提出的条件检测机制结合增强型特征可以很好的在目标消失又重现的过程中找回目标. 因此,综合前景概率图下的增强型特征与条件检测机制,使得本文的 STSF 算法具有很好的鲁棒性和有效性.

5 结论

在本文中我们提出一种新颖的基于目标显著性特征的可选择性跟踪和检测框架. 特征方面结合了前景概率图,增强了前景信息和背景的对比,突出了目标的显著性. 从长期跟踪过的角度出发,本文进一步提出了条件检测机制来辅助整体的算法实现精准鲁棒的跟踪. OTB 数据集上的实验也证明了算法是有效的.

参考文献

- [1] 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.
Yin Hong-Peng, Chen Bo, Chai Yi, et al. Vision based object detection and tracking: a review[J]. Acta Automatica Sinica, 2016, 42(10): 1466-1489. (in Chinese)
- [2] 闫雷兵,陆音,张业荣. 基于 TOA/TDOA 的一致性正交容积卡尔曼跟踪算法[J]. 电子学报,2018,46(8):1989-1996.

YAN Lei-bing, LU Yin, ZHANG Ye-rong. Unified orthogonal cubature Kalman tracking algorithm based on TOA/TDOA[J]. Acta Electronica Sinica, 2018, 46(8): 1989-1996. (in Chinese)

- [3] 盛琥,赵温波,张远. 基于转换量测高斯化的三坐标雷达跟踪算法[J]. 电子学报,2018,46(3):562-568.
SHENG Hu, ZHAO Wen-bo, ZHANG Yuan. 3D radar tracking algorithm with Gaussianized converted measurement[J]. Acta Electronica Sinica, 2018, 46(3): 562-568. (in Chinese)
- [4] 李凡,熊家军,李冰洋,等. 临近空间高超声速跳跃式滑翔目标跟踪模型[J]. 电子学报,2018,46(9):2212-2221.
LI Fan, XIONG Jia-jun, LI Bing-yang, et al. A tracking model for near space hypersonic jumping gliding target[J]. Acta Electronica Sinica, 2018, 46(9): 2212-2221. (in Chinese)
- [5] 李康,邵芳. 基于卷积神经网络的鲁棒高精度目标跟踪算法[J]. 电子学报,2018,46(9):2087-2093.
LI Kang, SHAO Fang. A robust and accurate object tracking algorithm based on convolutional neural network[J]. Acta Electronica Sinica, 2018, 46(9): 2087-2093. (in Chinese)
- [6] Wu Y, Lim J, Yang M H. Online object tracking: a benchmark[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. Portland: IEEE Press, 2013. 2411-2418.
- [7] Hong Z, Chen Z, Wang C, et al. Multi-store tracker (muster): A cognitive psychology inspired approach to object tracking[A]. Proceedings of the IEEE Conference on

- Computer Vision and Pattern Recognition [C]. Boston: IEEE Press, 2015. 749 – 758.
- [8] Zhang J, Ma S, Sclaroff S. MEEM: robust tracking via multiple experts using entropy minimization [A]. European Conference on Computer Vision [C]. Zurich: Springer, Cham, 2014. 188 – 203.
- [9] Song Y, Ma C, Wu X, et al. VITAL: Visual tracking via adversarial learning [J]. arXiv Preprint, arXiv: 1804.04273, 2018.
- [10] Lukežič A, Zajc L Č, Kristan M. Deformable parts correlation filters for robust visual tracking [J]. IEEE Transactions on Cybernetics, 2018, 48 (6): 1849 – 1861.
- [11] Bertinetto L, Valmadre J, Henriques J F, et al. Fully-convolutional Siamese networks for object tracking [A]. European Conference on Computer Vision [C]. Amsterdam: Springer, Cham, 2016. 850 – 865.
- [12] Galoogahi H K, Fagg A, Lucey S. Learning background-aware correlation filters for visual tracking [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. Venice: IEEE Press, 2017. 1144 – 1152.
- [13] Henriques J F, Rui C, Martins P, et al. High-speed tracking with kernelized correlation filters [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37 (3): 583 – 596.
- [14] Bertinetto L, Valmadre J, et al. Staple: Complementary learners for real-time tracking [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Las Vegas: IEEE Press, 2016. 1401 – 1409.
- [15] Ma C, Yang X, Zhang C, et al. Long-term correlation tracking [A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition [C]. Boston: IEEE Press, 2015. 5388 – 5396.
- [16] Li Y, Zhu J. A scale adaptive kernel correlation filter tracker with feature integration [A]. European Conference on Computer Vision [C]. Zurich: Springer, Cham, 2014. 254 – 265.
- [17] Danelljan M, Hager G, Shahbaz Khan F, et al. Convolutional features for correlation filter based visual tracking [A]. Proceedings of the IEEE International Conference on Computer Vision Workshops [C]. Santiago: IEEE Press, 2015. 58 – 66.
- [18] Danelljan M, Hager G, Shahbaz Khan F, et al. Learning spatially regularized correlation filters for visual tracking [A]. Proceedings of the IEEE International Conference on Computer Vision [C]. Santiago: IEEE Press, 2015. 4310 – 4318.

作者简介



丁新尧 男, 1996 年生于河南南阳. 本科毕业于华南理工大学, 现为华南理工大学电子与信息学院硕士研究生, 研究方向为计算机视觉与图像处理.
E-mail: scutdx@163.com



张鑫(通信作者) 女, 1981 年生于陕西西安. 现为华南理工大学电子与信息学院副教授, 主要研究方向为计算机视觉、图像处理和模式识别.
E-mail: eexinzhang@scut.edu.cn