

DOI:10.11830/ISSN.1000-5013.201601022



结合特征点匹配的在线目标跟踪算法

刘兴云, 戴声奎

(华侨大学 信息科学与工程学院, 福建 厦门 361021)

摘要: 提出一种结合特征点匹配的目标跟踪算法. 首先, 通过显著区域跟踪方法, 解决算法对初始化目标框大小敏感的问题, 提高样本选取质量, 并降低背景杂波对跟踪器的影响. 其次, 采用中值流法跟踪和特征点匹配相结合的方法估计目标的尺度变化, 并通过层级聚类方法剔除干扰点, 解决跟踪器漂移及目标平面旋转跟踪失败等问题. 最后, 提出一种简单的检测器自适应尺度快速搜索目标方法加快检测速度. 结果表明: 所提方法有效地提高了 TLD 目标跟踪算法的跟踪鲁棒性, 并在标准数据集上得到了很好的效果.

关键词: TLD; 目标跟踪; 显著性; 特征点匹配; 聚类

中图分类号: TP 391 **文献标志码:** A **文章编号:** 1000-5013(2018)03-0461-06

Online Target Tracking Algorithm Based on Feature Point Matching

LIU Xingyun, DAI Shengkui

(College of Information Science and Engineering, Huaqiao University, Xiamen 361021, China)

Abstract: A target tracking algorithm based on feature points matching has been proposed in this paper. Firstly, the method of tracking target in the saliency region can solve target size sensitivity while improving the quality of choosing samples and reducing the effect on the tracker caused by background clutter. Secondly, the target's scale can be estimated through median flow tracker and feature points matching and the interference point can be rejected by hierarchical clustering, based on these methods, tracker drifting and out-of-plane tracking failing can be resolved effectively. Finally, a simple fast search objectives method based on adaptive scale detector was proposed to accelerate detection speed. Experimental results demonstrate that the proposed algorithm can enhance the tracking robustness of TLD target tracking method effectively and obtain good results on standard data sets.

Keywords: TLD; target tracking; saliency; feature point matching; clustering

目标跟踪是计算机视觉研究领域中的一个关键问题,是视觉系统不可缺少的重要组成部分,已被广泛地应用于各个领域,如运动分析、智能监控、自动驾驶及人机交互等. 近年来,目标跟踪得到快速的发展,但是,在实际应用中依然面临多方面的挑战,如光照变化、目标形变、运动模糊、目标被遮挡或消失、复杂背景等. 为了解决这些问题,研究人员提出了各种解决方法^[1-4]. 虽然目标跟踪技术取得了很大的进展,但依然存在很多问题. 这些方法容易在目标外观迅速变化时产生跟踪漂移,且未考虑目标周围的背景信息,使目标在复杂的环境中跟踪质量显著降低. 判别式跟踪模型将目标跟踪问题视为二分类问题,通过分类器的学习将目标从背景中分离出来,并在线更新目标模板^[5-8]. 本文提出一种结合特征点匹配的在线目标跟踪方法.

收稿日期: 2016-01-12

通信作者: 戴声奎(1971-),男,副教授,博士,主要从事图像处理、模式识别的研究. E-mail:d. s. k@hqu. edu. cn.

基金项目: 福建省科技计划重点项目(2013H0030); 中央高校基本科研专项(JB-ZR1145)

1 TLD 跟踪算法

TLD 目标跟踪模型,如图 1 所示.算法主要由 4 个部分组成,即跟踪器、检测器、学习模块和综合模块.跟踪器利用前向、后向误差,对跟踪到的轨迹进行误差估计.如果误差大于一定的阈值,说明跟踪失败,从而选择可信度更高的轨迹.检测器采用级联分类器,由方差分类器、集合分类器和最近邻分类器组成,对当前视频中的所有扫描窗口进行检测,估计被跟踪目标的大致区域.学习模块根据输入的跟踪和检测结果生成新的正样本和负样本,更新在线模型,实现对检测器的更新.综合模块综合检测的结果和跟踪器的结果输出下一帧目标边界框的位置信息.

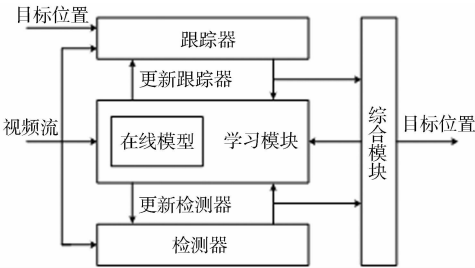


图 1 TLD 跟踪模型
Fig. 1 TLD tracking model

2 结合特征点匹配的跟踪方法

根据 TLD 跟踪算法存在对初始化目标框敏感,以及处理运动模糊、严重遮挡、形变及目标旋转等跟踪效果不佳的问题,提出一种结合特征点匹配的在线目标跟踪方法,具体可分为以下 3 点. 1) 计算跟踪目标的显著区域,提取最佳跟踪区域. 2) 结合光流法和特征点匹配进行跟踪,采用层级聚类方法剔除不稳定跟踪点. 3) 利用跟踪得到的尺度估计辅助检测器进行自适应尺度搜索,提高检测速度.

2.1 跟踪目标显著区域提取

文献[2]给出不同初始化目标框对跟踪算法的影响.当初始化目标框尺度较小时,TLD 跟踪算法表现出极好的性能.当初始化目标框较大时,目标包含了较多的背景信息,在跟踪过程中,背景的跟踪点对最后目标框的估计产生了影响,同时,生成的学习样本也含有越来越多的背景信息,形成误差累积,最终导致跟踪失败.背景对跟踪的影响,如图 2 所示.为了解决上述问题,采用目标显著区域跟踪法,通过提取待跟踪目标的显著区域消除 TLD 跟踪算法受背景杂波的影响,且小目标受目标的形变影响更小,从而提高跟踪的鲁棒性.

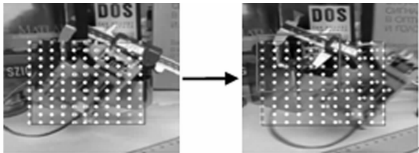


图 2 背景对跟踪的影响
Fig. 2 Influence of background on tracking

采用显著性提取方法^[9],目标显著性区域的提取,如图 3 所示,具体有以下 2 点.

- 1) 对跟踪目标 Box 进行 $8\text{ px} \times 8\text{ px}$ 的无重叠分割,计算各个块的显著性,并提取显著性值高的块作为候选跟踪区域,如图 3(a),3(b)所示.
- 2) 如果显著块少于阈值 $\text{Thr}_{\text{patch}}$,则用原来的目标框进行跟踪;否则,以目标框的中点为中心,步长为 8,根据跟踪目标的长宽比例向外扩展.计算当前扩展区域所含的显著块的面积与扩展区域总面积的比值 S_{saliency} .当比值小于 $\text{Thr}_{\text{score}}$ 时,停止搜索.此时,得到新的跟踪目标框,如图 3(c)所示.

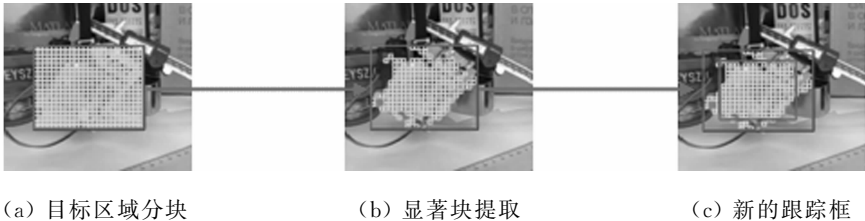


图 3 目标显著性区域提取
Fig. 3 Saliency region extraction of tracking target

2.2 融合光流法和特征点匹配的跟踪器

TLD 跟踪器采用中值光流法进行跟踪,在目标发生平面旋转时,会产生跟踪漂移现象,导致跟踪失败.其主要原因是当目标遇到上述跟踪困难因素时,不满足光流法的假设条件,导致跟踪失败.更重要的是,TLD 跟踪器固定不变易产生相同的错误,当跟踪失败时,检测器初始化的跟踪目标不一定准确,易产生误差累积,导致跟踪失败.因此,采用特征点匹配法提高跟踪性能.首先,提取待跟踪区域两倍范围

内的特征点, 将待跟踪目标区域的特征点标记为前景特征点, 剩下的标记为背景特征点并保存, 对前景特征点分别进行前、后向光流法跟踪和特征点匹配, 将跟踪到的点和匹配到的点合并, 剔除与背景点相匹配的点, 然后, 采用层级聚类的方法剔除不稳定的跟踪点, 并估计目标的位置及大小。

2.2.1 结合中值光流法和特征点匹配的跟踪器 为了解决上述问题, 采用 FAST 特征点检测快速检测特征点, 采用具有旋转不变性的 BRISK 描述子对特征点进行描述, 对特征点进行中值流法跟踪和特征点匹配共同估计目标的位置和尺度。

对跟踪器进行初始化, 对初始化目标框两倍范围内的区域进行 FAST 特征点检测, 计算特征点的 BRISK 描述子并保存, 可得

$$P = \{(l_1, f_1), (l_2, f_2), \dots, (l_n, f_n)\}. \quad (1)$$

式(1)中: $l_{1:n}$ 为所提取的特征点位置; $f_{1:n}$ 为所提取特征点的二进制描述子. 将特征点划分为前景特征点 $P_{fg} = (l_1, f_1), \dots, (l_{n_1}, f_{n_1})$ 及背景特征点 $P_{bg} = (l_1, f_1), \dots, (l_{n-n_1}, f_{n-n_1})$. 从第二帧开始对上一帧跟踪到的目标区域两倍范围内的区域进行特征点检测, 可得

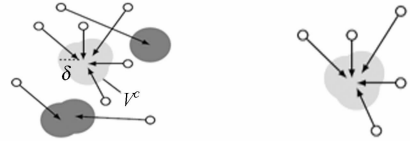
$$M = \{(r_1, f_1), (r_2, f_2), \dots, (r_{n_2}, f_{n_2})\}. \quad (2)$$

式(2)中: $r_{1:n_2}$ 是所检测到的特征点位置. 由于 BRISK 描述子是二进制描述子, 因此, 采用两者的汉明距离判断特征点的相似性, 即

$$H = \sum_{i=1}^d \text{XOR}(f_i^P, f_i^M). \quad (3)$$

通过最近邻距离和次邻近距离的比值 θ 确定匹配有效性, 得到有效的匹配点, 剔除与背景相匹配的特征点, 得到跟踪目标的特征匹配点 M_{fg} . 在匹配的同时, 对 P_{fg} 特征点进行前、后向光流法跟踪, 得到 FB_error 小于 Thr_{FB} 的跟踪点 T_{fg} , 融合匹配得到的特征点 M_{fg} 和跟踪得到的特征点 T_{fg} , 以及它们的并集点 P_{fuse} . 估计特征点之间的相对位移, 采用中值估计法估计目标的位置变化、尺度变化及角度变化。

2.2.2 结合层级聚类的干扰点剔除 为了精确地确定目标的位置, 采用中心点投票法^[10]剔除错误跟踪点, 估计目标所在的位置, 并得到下一帧的待跟踪点. 由于跟踪和匹配过程不可避免地产生错误, 这些点严重影响了目标中心的位置估计. 采用层级聚类法^[11]提取最大类, 并剔除干扰点. 将跟踪的特征点进行尺度和角度变换, 由于跟踪过程中会产生误差, 所产生的新的坐标中心也会存在误差, 因此, 引入投票机制剔除干扰点. 干扰点的剔除, 如图 4 所示. 对最终变换的坐标点进行自下而上的层级聚类, 基于大部分特征点的偏差比较小的理论, 所跟踪到的特征点大部分围绕在真正目标中心附近, 聚集在一个较小的区域, 通过聚类提取数目较多且占总体比例较高的类, 并通过



(a) 跟踪到的候选特征点 (b) 聚类后剩下的特征点
图 4 干扰点的剔除

Fig. 4 Eliminate noise points

$$\text{center} = \frac{1}{k} \sum_{i=1}^k V_i^c \quad (4)$$

估计出目标的准确中心点. 式(4)中: k 为聚类点数; V_i^c 为 V^c 中的点. 结合所得到的目标尺度即可确定目标的精确位置和大小。

2.3 检测器自适应尺度搜索

由于跟踪器采用匹配的方法可以估计出目标尺度, 且相邻帧的目标尺度变化不大, 因此, 为了提高检测器的检测速度及跟踪算法的实时性, 根据跟踪器的尺度估计, 设计一种自适应尺度检测方法, 具体有以下 4 点. 1) 存储跟踪器最近 10 帧的目标尺度变化 $S = [s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9, s_{10}]$. 2) 计算尺度变化 S 的中值 S_{Med} , 并计算 S_{Med} 与初始化尺度间的关系. 3) 通过映射关系找到对应的网格化尺度 s_{grid} , 用检测器对 $(s_{\text{grid}} - 6) \sim (s_{\text{grid}} + 6)$ 尺度进行检测. 4) 如果跟踪失败, 则对当前帧进行全部尺度检测。

3 实验结果与比较

为了验证文中算法的有效性, 在 25 个测试视频上对提出的算法进行对比验证. 这些视频来自于文

献[2]的测试库,所有视频的 Ground truth 文件均来自该测试库.这些视频序列包含了目标跟踪中所遇到的各种困难因素,包括光照变化、尺度变化、严重遮挡、背景杂波、运动模糊、快速运动、平面旋转及非刚性形变等,具有较高的代表性.为了保证实验的可信度,挑选了公认目标跟踪效果较好的 6 种跟踪算法,并采用作者提供的公开源代码或可执行文件,这些算法包括 CSK^[12],STC^[13],MIL^[5],Struck^[7],CT^[8],TLD^[6].

开发平台为 Visual Studio 2012,Opencv 2.4.6,Eigen 3.2,编程实现上述算法,并在 Intel(R) Core (TM) i5-4590 CPU@3.30 Hz with 4 GB RAM 的计算机上进行测试.实验参数设置:Thr_{batch} = 50; Thr_{score} = 0.7; Thr_{FB} = 10; θ = 0.8; d = 512. 由于算法在跟踪时候的初始化目标框进行了缩放,所以在算法评估时按照原始的缩放比例进行缩放.采用的默认准则为:对具有跟踪失败的算法采用前一帧的目标框来替代,并使用公开代码中的默认参数,在实验过程中,保持参数不变.

3.1 定性分析

部分跟踪结果截图,如图 5 所示.所选的测试视频中,carscale 视频序列主要包含目标尺度的大幅度变化及目标遮挡现象,如图 5(a)~(d)所示.由图 5(a)~(d)可知:只有文中算法能够准确地估计目标的尺度变化,在 carscale 视频序列中,跟踪成功率超过 50%的算法只有 TLD 算法和文中算法,且文中算法的跟踪成功率比 TLD 算法高 20%.这主要是因为文中算法通过特征点匹配,提高目标抗遮挡的能力,通过引入自下而上的层级聚类剔除不可靠的跟踪点,准确地估计目标的中心位置及尺度信息,获得较好的跟踪效果.



图 5 部分跟踪结果截图

Fig. 5 Screenshot for part tracking results

由图 5(e)~(h)可知:目标存在严重的旋转,目标出现从最开始的正脸到侧脸,再到另一边的侧脸;到 #95 帧时,只有文中算法可以准确跟踪到目标的位置及大小,最后跟踪成功率超过 50%的算法只有文中算法.这是因为原始的 TLD 算法采用光流法进行目标跟踪,当目标发生旋转时,光流法的假设条件不再满足,跟踪失败.文中采用特征点检测及 BRISK 描述子,具有抗目标旋转的性能,可以有效地跟

踪旋转目标. woman 序列含有非刚性形变、严重遮挡及运动模糊等, 给目标跟踪带来很大的跟踪困难. 由图 5(i)~(l)可知: 当跟踪的目标被汽车遮挡后, MIL, CT, TLD 算法均发生漂移, 而到 # 282 帧时, STC, CSK 算法也无法正确地跟踪目标; 到 # 564 帧时, 目标发生了镜头抖动、大小变化、形变及运动模糊时, 只有文中算法可以有效地跟踪目标. 当目标跟踪场景中出现遮挡及相似目标时, CSK, Struck 算法出现跟踪漂移现象. 由图 5(m)~(p)可知: 到 # 381 帧时, 由于目标经过瓶子时出现遮挡, MIL, CT, TLD 算法均发生偏移; 到 # 928 帧时, 由于目标反复遇到严重遮挡, CSK, STC 算法均没有处理严重遮挡的能力, 跟踪失败且具有不可恢复能力, 而 Struck 算法遇到相似目标时也发生了误跟踪, 只有文中算法可以较好地应对这些困难, 获得了第一的跟踪成功率和距离精度, 都接近 100%. 主要原因在于为了防止目标遇到严重遮挡, 防止跟踪器初始化时发生误跟踪的情况, 文中算法在第一帧初始化目标的时候就保留了目标及其周围的特征点, 在目标遇到遮挡时, 检测器重新对跟踪器进行初始化, 只有特征点匹配相似度很高的时候才进行重新初始化, 这样可以避免检测器的误初始化, 防止误跟踪, 同时, 可以准确地估计出目标的尺度信息, 精确地跟踪目标.

3.2 定量分析

采用跟踪成功率进行定量分析, 跟踪成功率定义为 $SC = \frac{\text{area}(R_T \cap R_G)}{\text{area}(R_T \cup R_G)}$. 式中: R_T, R_G 分别为跟踪算法和人工手动标注的目标状态; area 为目标跟踪状态的面积. 当 $SC > 0.5$ 时, 表示目标跟踪成功. 跟踪结果的成功率, 如表 1 所示. 由表 1 可知: 文中算法在多数视频中取得较好的成绩, 相对于目前主流的跟踪算法, 文中算法在测试序列上的平均成功率最优, 这表明算法是合理有效的.

表 1 跟踪结果的成功率
Tab. 1 Tracking success rate

序列	SC 值						
	CSK 算法	STC 算法	MIL 算法	Struck 算法	CT 算法	TLD 算法	文中算法
boy	0.81	0.64	0.310	0.97	0.70	0.81	0.95
carscale	0.44	0.46	0.440	0.41	0.41	0.50	0.73
coke	0.73	0.05	0.110	0.94	0.09	0.51	0.84
couple	0.08	0.08	0.670	0.54	0.68	1.00	0.60
deer	0.98	0.04	0.113	0.98	0.04	0.73	0.95
dog1	0.60	0.56	0.350	0.61	0.46	0.66	0.99
doll	0.33	0.10	0.350	0.47	0.54	0.61	0.98
dudek	0.86	0.68	0.820	0.91	0.83	0.80	0.96
faceocc1	1.00	0.23	0.810	0.98	0.85	0.79	0.96
fish	0.04	0.37	0.370	1.00	0.86	0.96	1.00
fleetface	0.66	0.38	0.480	0.67	0.55	0.49	0.61
football	0.73	0.60	0.710	0.66	0.78	0.41	0.69
football1	0.43	0.39	0.820	0.81	0.08	0.39	0.52
freeman1	0.13	0.13	0.130	0.19	0.10	0.21	0.62
jumping	0.04	0.04	0.580	0.95	0.00	0.84	0.97
liquor	0.27	0.25	0.200	0.40	0.20	0.56	0.96
matrix	0.01	0.02	0.110	0.12	0.02	0.06	0.10
mountainbike	1.00	0.61	0.540	0.95	0.16	0.26	0.93
shaking	0.83	0.82	0.190	0.13	0.04	0.38	0.73
singer1	0.27	0.48	0.270	0.29	0.20	0.99	0.90
skating1	0.36	0.21	0.100	0.29	0.05	0.21	0.59
subway	0.21	0.22	0.800	0.89	0.77	0.22	0.65
suv	0.57	0.49	0.130	0.57	0.23	0.83	0.90
walking2	0.40	0.57	0.380	0.42	0.27	0.39	0.76
woman	0.24	0.21	0.180	0.93	0.15	0.16	0.78
平均值	0.48	0.34	0.390	0.64	0.36	0.55	0.78

3.3 计算时间分析

文中所用的目标显著区域提取方法只有在目标初始化时进行一次提取,之后将进行正常跟踪,基本不会影响跟踪系统的时效性.由于算法采用了局部特征点匹配,在一定程度上降低跟踪算法的实时性,但提出了检测器自适应尺度搜索,在不影响跟踪精度的情况下,弥补了特征点匹配所耗费的时间.在实验测试中,25 个视频序列的平均运行速度为 $24.6 \text{ 帧} \cdot \text{s}^{-1}$,平均跟踪成功率达到 79.2% ,平均距离精度达到 86.7% .而在没有采用检测器自适应尺度搜索时,对测试序列的平均运行速度为 $16.8 \text{ 帧} \cdot \text{s}^{-1}$,平均跟踪成功率为 79.8% ,平均距离精度达到 87.3% .因此,提出的自适应尺度搜索方法在不影响跟踪性能的前提下,提高了跟踪实时性.

4 结束语

提出一种结合特征点匹配的在线目标跟踪方法,在原始 TLD 基础上对其改进.实验结果表明:提出的算法更加优秀,平均跟踪成功率超过了原始 TLD 算法及当前的主流目标跟踪算法,验证了文中算法的有效性.改进后的 TLD 算法对不同场景下、不同类型的视频目标都有很好地跟踪效果,具有更好的场景适应性.

参考文献:

- [1] YILMAZ A, JAVED O, SHAH M. Object tracking: A survey[J]. ACM Computing Surveys, 2006, 38(4): 13. DOI: 10.1145/1177352.1177355.
- [2] WU Yi, LIM J, YANG M H. Online object tracking: A benchmark[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Oregon: IEEE Press, 2013: 2411-2418. DOI: 10.1109/CVPR.2013.312.
- [3] ADAM A, RIVIN E, SHIMSHONI I. Robust fragments-based tracking using the integral histogram[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York: IEEE Press, 2006: 798-805. DOI: 10.1109/CVPR.2006.256.
- [4] XU Cheng, LI Nijun, ZHANG Suofei, *et al.* Robust visual tracking with SIFT features and fragments based on particle swarm optimization[J]. Circuits, Systems, and Signal Processing, 2014, 33(5): 1507-1526. DOI: 10.1007/s00034-013-9713-1.
- [5] BABENKO B, YANG M H, BELONGIE S. Robust object tracking with online multiple instance learning[J]. Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8): 1619-1632. DOI: 10.1109/TPAMI.2010.226.
- [6] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection[J]. Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422.
- [7] HARE S, SAFFARI A, TORR P H S. Struck: Structured output tracking with kernels[C]// Proceedings of the IEEE International Conference on Computer Vision, Barcelona: IEEE Press, 2011: 263-270. DOI: 10.1109/ICCV.2011.6126251.
- [8] ZHANG Kaihua, ZHANG Lei, YANG M H. Fast compressive tracking[J]. Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(10): 2002-2015.
- [9] LEE D Y, SIM J Y, KIM C S. Visual tracking using pertinent patch selection and masking[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus: IEEE Press, 2014: 3486-3493. DOI: 10.1109/CVPR.2014.446.
- [10] NEBEHAY G, PFLUGFELDER R. Consensus-based matching and tracking of keypoints for object tracking[C]// IEEE Winter Conference on Application of Computer Vision, Steamboat: IEEE Press, 2014: 862-869. DOI: 10.1109/WACV.2014.6836013.
- [11] XU Rui, WUNSCH D. Survey of clustering algorithms[J]. Neural Networks, 2005, 16(3): 645-678.
- [12] HENRIQUES J F, RUI C, MARTINS P, *et al.* Exploiting the circulant structure of tracking-by-detection with kernels[J]. Lecture Notes in Computer Science, 2012, 7575(1): 702-715.
- [13] ZHANG Kaihua, ZHANG Lei, LIU Qingshan, *et al.* Fast visual tracking via dense spatio-temporal context learning [C]// European Conference on Computer Vision, Zurich: Springer International Publishing, 2014: 127-141.