红外图像实时在线离线跟踪算法研究

张希铭, 王民钢, 曹琳

(西北工业大学 航天学院, 陕西 西安 710072)

摘 要:大多数基于目标检测的红外图像目标跟踪算法采取基于时空一致性在线模型更新策略。然而,当所跟踪的目标发生形变、快速运动和受到遮挡时,在线模型更新过程会受到不同程度的干扰而导致目标跟踪失败。基于孪生网络的离线模型跟踪策略则能够在目标发生扰动的情况下保持其外观模型的不变性。然而,在跟踪速度上与在线模型更新策略差距较大。提出了目标跟踪过程中的跟踪错误检测方法将在线和离线目标模型更新方法相结合,该检测方法通过基于联合响应图的离散度测量来联合 2 类模型更新方法,并能根据当前目标跟踪状态自动在 2 种模型更新方法中切换,有效地解决了跟踪算法实时性与鲁棒性的平衡问题。所提出算法在 VOT-TIR-2015 数据库的实验结果显示相比原有算法 Staple 和 SiamFC 在跟踪成功率上分别提高 3.3%和 3.6%,在跟踪精度上分别提高 3.8%和 5%,同时保证跟踪的实时性。

关 键 词:红外目标跟踪;Staple;孪生网络;错误跟踪检测

中图分类号:TP391 **文献标志码**:A **文章编号**:1000-2758(2018)06-1052-07

随着红外摄像头在成像质量、分辨率等属性上得到增强且成本逐渐降低、尺寸日渐缩小,红外图像应用的领域更加广泛。最初,红外图像主要应用于军用领域来实现复杂背景下的小目标(点目标)跟踪。随着成像科技的进步,红外图像应用的领域也逐步得到了拓展,例如车辆与监控系统。除此之外,搭载的方式也不仅限于固定的平台,搭载红外成像系统的无人机平台主要用于侦查与救援。红外成像主要的优势是能够在完全黑暗的情况下使用,并对光照变化有较强的鲁棒性。

红外目标的跟踪问题面临的挑战主要包括以下 3 点:①连续性:视频流必须以一定顺序进行处理;②即时性:跟踪算法在跟踪过程中不允许对目标真实框体重新初始化;③未知性:在目标跟踪前不允许使用对于目标外观的预构建模型。

现阶段在 VOT-VIR-2015(红外图像)数据库[1]中公布的针对红外目标跟踪的算法共有 24 种。其中包括:利用似物性实例来实现检测再跟踪的 EBT、基于 MeanShift 中值漂移及其拓展的跟踪算法 ASMS、基于在线更新级联分类器方法实现的 CMIL、基于检测再跟踪的学习方法 sPST 和基于部件的跟踪算法 LDP。基于辨别式方法的跟踪算法在学术领域研究甚广,其中最常用的是基于结构支持向量机的 Struck,还有采用相关滤波器来获取目标外观模型的方法,其中包括利用相关滤波器获取单一外观模型的 sKCF 和利用多模板来构建外观模型的 KCFv2。

由于在线模型更新办法依靠其时空一致特性来实现,因此在红外目标跟踪过程中对于目标自身发生形变、受到遮挡干扰和快速运动模糊的情况下处理效果不好,例如 3 种在线更新算法(Struck[2], DSST[3]和 Staple[4])在所跟踪红外目标遭遇人体遮挡的情况下会发生跟踪漂移。

基于离线跟踪的目标跟踪算法的外观模型是通过离线训练得出的,因此在跟踪的模型更新过程中对于干扰具有一定的鲁棒性。相比于在线更新算法,离线更新算法将跟踪问题作为类学习问题来处理。例如孪生网络 SiamFC[5],简化深度网络结构,利用 5 层卷积网络在模板图像和搜索图像上进行操作,并利用相似性度量函数将 2 类图像矩阵联合构

收稿日期:2018-01-08 **基金项目**:国家自然科学基金(61703337)与上海航天科技创新基金(SAST2017-082)资助

作者简介:张希铭(1989—),西北工业大学博士研究生,主要从事目标检测与跟踪及深度学习方法研究。

成得分矩阵,再根据得分矩阵结果来估计下一帧目标位置。离线更新模型在只给出第一帧的初始信息的情况下,就可以在不经历在线更新的情况下得到比较优异的结果。然而,由于没有在线更新的过程,这类算法容易错误跟踪与第一帧中目标的外观相似的伪目标。当目标跟踪过程中经历相似目标时,目标跟踪器会发生漂移而导致跟踪失败。

因此,目前的红外目标跟踪算法,在线更新和离线模型更新策略之间的选择是一种博弈,一方面,离线训练的外观模型对于模型更新具有鲁棒性,然而并不能处理好特定的外观模型变化,在长期跟踪过程中无法保证良好的跟踪效果。另一方面,具有高度适应性的在线模型更新算法可能会由于受到各类外界的干扰而发生漂移。

本文提出了基于在线和离线模板更新相结合的红外目标跟踪算法,利用 2 种方法的优势弥补不足,通过 2 种策略的结合来得到更好的跟踪效果,其中在线更新部分采用 Staple 算法,离线更新部分采用孪生网络 SiamFC,并提出基于联合响应图的离散度测量方法来进行错误跟踪检测以实现算法在 2 种模式之间的切换。本文在 VOT-VIR-2015 数据库评估了所提出的算法。相比于原算法 Staple 和 SiamFC 在跟踪成功率方面提高了 3.3% 和 3.6%,在跟踪精度上分别提高了 3.8% 和 5%,同时保持了跟踪速度平均 46.7 frames/s,具有较高的实时性。

1　算法流程

本文提出算法为红外在线离线跟踪算法,TCOOT(thermal collaborative online and offline tracker)。算法首先通过提取 HOG 和颜色特征分别计算模板响应和直方图响应来估计平移量并产生联合响应图,再将联合响应图传送至错误跟踪检测器中计算离散度测量值与自适应切换阈值,以捕捉在线模板更新可能出现错误的关键点,再通过判断 2 组值之间的大小关系,决定切换至在线还是离线跟踪策略来进行平移与尺度估计,循环执行以上步骤,直至视频处理结束。红外在线离线跟踪算法的整体流程图,如图 1 所示。

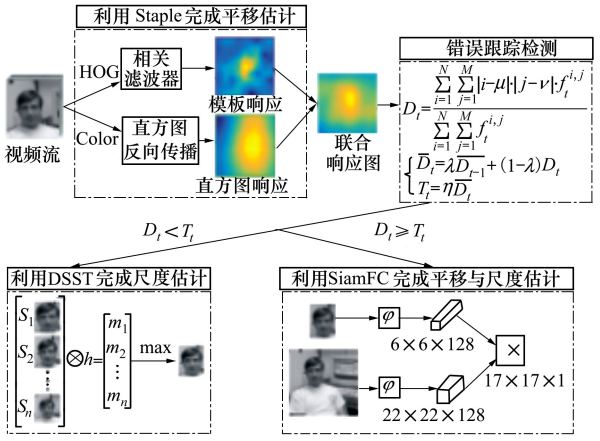


图 1　红外在线离线跟踪算法流程图

2　实时在线离线目标跟踪

红外在线离线跟踪算法(TCOOT)主要的组成部分包括:基于相关滤波器的在线模板更新跟踪器,基于孪生网络的离散跟踪器和错误跟踪检测器。

2.1　基于相关滤波器在线模板更新跟踪器

传统基于相关滤波器的在线更新跟踪算法采用目标的空间结构来完成模板建模(例如 HOG),在跟踪目标发生形变时表现差强人意。采用颜色统计(例如灰度直方图)的方法虽然对于形变有一定鲁棒性,但是光照的变化对于其外观模型的影响较大。通过结合 HOG 特征与灰度直方图的在线更新模型算法 Staple 在多种数据库上都实现了非常好的效果,对于目标自身形变和光照变化适应性较强,并且保证了目标跟踪的实时性,跟踪速度保持在 81 frames/s 。

本文采用 Staple 作为在线模型更新方法,因为相比较于其他基于相关滤波器的目标跟踪算法,Staple 在模型的复杂度和算法的执行效率上保持了很好的平衡。在应用 Staple 作为在线模型更新工具的情况下可以实现目标外观模型的实时处理。Staple 结合模板建模和直方图的联合响应图在处理复杂跟踪情境时可以通过设计错误跟踪检测器来完成判别与切换,为本文算法的实现提供先决条件。

2.2　基于孪生网络的离线跟踪器

现阶段有很多利用孪生网络为基础框架的跟踪模型,在没有使用任何在线模型更新机制的情况下

均取得了很好的效果。孪生网络采用深度学习框架,并通过 IILSVRC15 数据库中的图像进行训练,从而获得网络模型的参数,用于提取目标图像特征,相比于在线更新跟踪算法,基于离线更新的孪生网络对于模型的更新过程中的干扰的适应性更强。除此之外,孪生网络 SiamFC 利用第一帧目标所在位置作为模板图像,将其他帧中目标所在区域构成搜索图像,在搜索图像上利用模板图像进行相似性匹配以确定目标的最终位置,算法简单,参数固定。然而,需要注意的是基于孪生网络的离线跟踪算法在经历相似目标或者第一帧的目标外观与其他的帧相差迥异时会出现跟踪失败的情况。

通过分析在线和离线模板更新的优势和不足,本文提出错误跟踪检测器在跟踪过程中实现 2 种跟踪器的过程判定与自动切换以保证目标跟踪算法的准确鲁棒跟踪。

2.3 错误跟踪检测器

错误跟踪检测器用于检测在跟踪过程中错误发生的关键点,本文采用针对联合响应图的离散度测量方法来实现错误跟踪检测。一般情况下,联合响应图在目标所在位置达到峰值,并且从离散中心到边界很快得到抑制。然而在目标经历严重遮挡,形变和快速运动时,联合响应图的峰值并没有那么高,并且联合响应图抑制的速度很慢,如图 2 所示。其中图 2a) 表示 TCOOT 在视频流 Birds 上的跟踪结果;图 2b) 表示 Staple 算法在 Birds 对应帧上的联合响应图。在第 83 帧,Staple 由于遮挡原因跟丢了目标,因此响应图是发散的。在这种情况下,SiamFC 被自动激活并搜索目标。在第 96 帧,响应图上最高值回到中心,继续用 Staple 来跟踪目标。

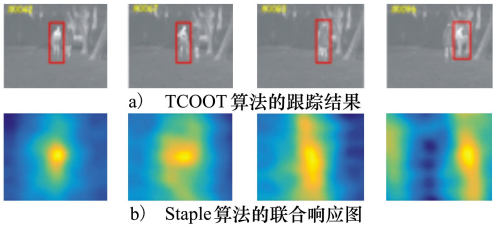


图 2 跟踪结果与联合响应图关系示意图

结合联合响应图这个重要特性,本文提出了离散度测量方法,并利用其判定跟踪目标的当前状态并预防可能的跟踪错误。

首先计算联合响应图的峰值位置。令 $\boldsymbol{f}_t$ 为 $N \times M$ 的矩阵表示第 t 帧的联合响应图。峰值 $[\mu, \upsilon]$ 的位置通过优化(1) 式得到:

$$[\mu, \upsilon]^{\mathrm{T}} = \arg\max_{i,j} f_t^{i,j} \tag{1}$$

式中,$f_t^{i,j}$ 对应 $\boldsymbol{f}_t$ 中第 i 行的第 j 个元素。

对于 $\boldsymbol{f}_t$ 的离散度测量如(2) 式所示:

$$D_t = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} |i-\mu| \cdot |j-\upsilon| \cdot f_t^{i,j}}{\sum_{i=1}^{N}\sum_{j=1}^{M} f_t^{i,j}} \tag{2}$$

之后,通过公式(3) 来计算每一帧对应的离散度测量值 D_t 的自适应切换阈值 T_t。

$$\begin{cases} \bar{D}_t = \lambda \bar{D}_{t-1} + (1-\lambda) D_t \\ T_t = \eta \bar{D}_t \end{cases} \tag{3}$$

式中,$0 < \lambda < 1$ 是遗忘因子,(3) 式通过遗忘因子来调整离散度测量值的敏感度,值越大敏感度越低,相反则越高。$\eta > 1$ 是放大因子,(3) 式通过放大因子来平衡离散度测量值与自适应阈值的相对位置,值越大越容易触发算法切换,相反则越难。$\bar{D}_t$ 是所更新的 D_t 的均值。视频流 Birds 在跟踪过程中所有帧对应的离散度测量值 D_t 和自适应切换阈值 T_t ,如图 3 所示。由于在第 82 帧出现了严重的遮挡,离散度测量值激增到一个高值,因此推测 Staple 在此处会出现跟踪错误可能,算法自动切换至 SiamFC,且在第 89 帧成功地跟踪目标。在算法执行过程中,离散度测量值与自适应切换阈值由求解(3) 式获得,主要受到跟踪图像帧的联合响应图的影响,跟踪过程开始前 10 帧,错误检测机制关闭,等待联合响应图稳定后,跟踪过程完成 10 帧后自动开启。

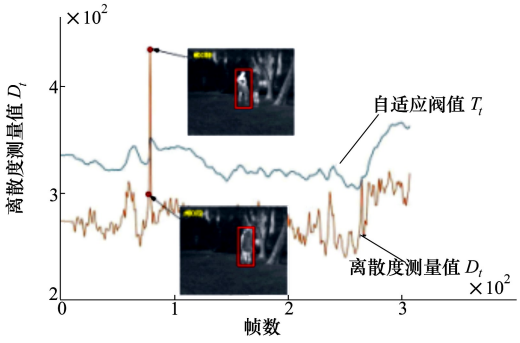


图 3 视频流 Birds 的离散度测量值和自适应切换阈值示意图

3 实验结果与分析

本文采用 VOT-TIR-2015 红外目标跟踪数据库

中的 20 个视频流对本文提出的 TCOOT 算法进行验证,实验过程中利用 TCOOT 分别与 Staple,SiamFC 和 DSST 3 种独立的算法进行对比。首先介绍该算法的完成细节与参数配置。接着,利用数据库中的红外目标跟踪视频流分别在距离精确度,覆盖成功率和不同情境下算法的表现对 4 种算法进行评估。

3.1　完成细节

TCOOT 算法采用 Staple 和 SiamFC 中模型参数默认的参数配置。在跟踪的过程中,主要调整错误跟踪检测器的参数。算法的仿真对比在数据库 VOT-TIR-2015 中进行,数据库中的视频流在目标运动过程中受到遮挡、旋转、形变、运动模糊、背景干扰和光照变化等干扰,可以有效地验证跟踪算法的精确度与鲁棒性。

(3)式中的遗忘因子 λ 和放大因子 η,在仿真实验中初始化为 $\lambda=0.9$ 和 $\eta=1.2$。错误跟踪检测器从第 10 帧开始工作,以保证 Staple 的联合响应图在视频流前端的稳定性。所有的参数在实验过程中保持不变。系统环境为 Intel i5 处理器 3.5GHz,8G 内存,NVidia GeForce GTX1080Ti GPU。算法运行环境为 Matlab 与 MatConvNet[6]。

算法 1　红外图像实时在线离线跟踪算法

输入　初始化目标位置 $\boldsymbol{X}_0$,自适应切换阈值 T_0

输出　估计目标位置信息 $\boldsymbol{X}_t=(x_t,y_t,s_t)$

算法开始:

1) 根据第 t 帧的目标位置信息$(x_{t-1},y_{t-1},s_{t-1})$来确定搜索框的位置与尺度;

2) 计算对应响应图 $\boldsymbol{F}_t$ 并利用 Staple 估计目标新位置(x_t,y_t);

3) 计算离散度测量值 D_t 与 $\boldsymbol{F}_t$ 自适应切换阈值 T_t;

4) 如果 $D_t < T_t$

利用 DSST 来估计新的尺度 s_t;

利用 Staple 来更新模版和直方图模型;

利用 DSST 来更新一维尺度相关滤波器。

如果 $D_t \geqslant T_t$

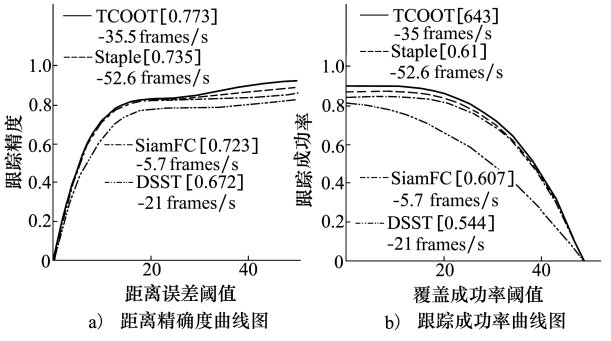
重新利用孪生网络离线估计位置和尺度;

5) 循环处理至视频流跟踪结束

算法结束

3.2　跟踪定量评估

定量评估的主要评估指标分别为距离精确度和覆盖成功率。采用一次性评估方法将 TCOOT 算法与其他 3 种算法作对比。其中 Staple 和 DSST 是基于相关滤波器的在线更新算法。SiamFC 是基于孪生网络的离线更新算法。不同算法在一次性评估的距离精确度和覆盖成功率曲线图如图 4 所示。其中距离精确度是利用视频流中所有帧的估计目标位置与真实目标位置的中心位置平均欧式距离测度获得的。覆盖成功率是由估计跟踪目标的图像框与真实目标图像框的覆盖比获得的,假设估计跟踪目标图像框为 r_t,真实目标图像框为 r_0,那么覆盖成功率为 $\delta_{\mathrm{SUC}}=\dfrac{|r_t\cap r_0|}{|r_t\cup r_0|}$,其中 $\cap$ 和 $\cup$ 表示交集与并集,$|\cdot|$ 表示区域内图像帧的数量。评估结果按照 AUC (area under curve)方法进行排名,并依次排列在标注框中,图 4a)中横坐标表示估计图像框与真实图像框中心误差,纵坐标表示跟踪精确度;图 4b)中横坐标表示图像框覆盖率的阈值,纵坐标表示覆盖成功率。总体仿真实验定量分析结果如表 1 所示,TCOOT 算法在精确度和成功率上皆排在其他 3 种对比算法之前,虽然 TCOOT 是由 Staple 和 SiamFC 2 种算法结合而成的,但是在精确度和成功率上分别比 Staple 和 SiamFC 提高了 3.3%和 3.6%,在跟踪精度上分别提高了 3.8%和 5%,算法执行速度为 46.7 frames/s,由于其结合在线模型更新与离线模型跟踪策略,因此在跟踪速度上介于 Staple 与 SiamFC 之间,满足实时性要求。



a) 距离精确度曲线图　　b) 跟踪成功率曲线图

图 4　采用一次评估的距离精确度和跟踪成功率曲线图

表 1　总体仿真实验定量分析结果

跟踪器	精确度	成功率	算法实时性/(frames · s^{-1})
TCOOT	0.773	0.643	46.7
Staple	0.735	0.61	52.6
SiamFC	0.723	0.607	5.7
DSST	0.672	0.544	21

3.3 不同情境下算法评估

VOT-TIR-2015 数据库的 20 个图像视频流中包含 11 种不同的情境。其中主要分析 6 种不同的情境,包括自身形变、遮挡、面内旋转、离面旋转、复杂背景和光照变化。6 种不同情境的影响下各算法跟踪成功率的曲线图如图 5 所示,其中横纵坐标的表示与图 4b)中的表示相同。从图 5 可以看出基于在线更新的相关滤波器 Staple,DSST 在目标出现个体形变,光照变化和离面旋转时表现良好。基于离线的孪生网络在目标出现面内旋转,背景复杂,遮挡时表现优异。本文算法针对在线离线模型更新方式取长补短,利用跟踪错误检测机制在 2 种策略中切换,从而完成跟踪过程,因此本文提出 TCOOT 跟踪算法在 6 种约束上除离面旋转外均表现最优。

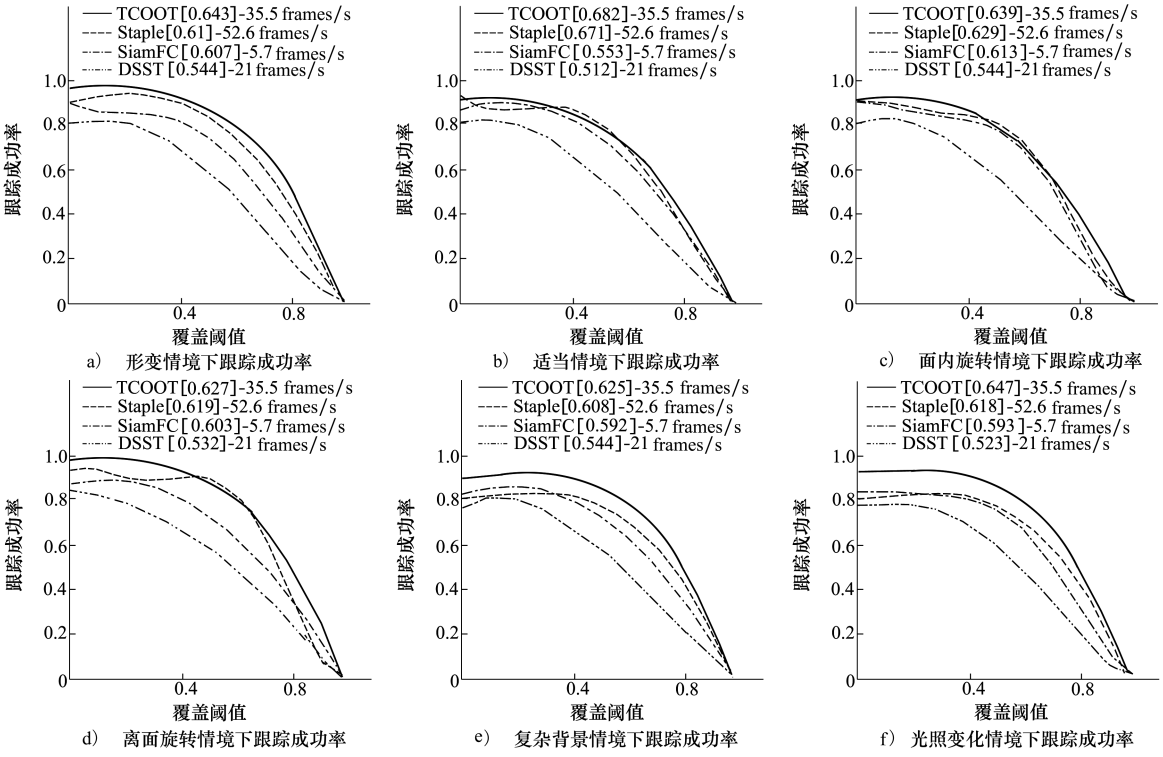


图 5 TCOOT 在 VOT-TIR-2015 上不同情境约束的成功率分析

3.4 跟踪定性评估

TCOOT 算法在跟踪的定性评估方面如图 6 所示。针对不同的挑战情境选取了 4 种红外视频流如遮挡分别如图 6a)~d)所示,TCOOT 在这些视频流中都表现出了很强的鲁棒性。

视频流四旋翼无人机中的红外目标的快速移动,视频流隐蔽中的红外目标发生遮挡,由于目标可视信息的不连续,Staple 在跟踪过程中更加容易丢失目标。在拥有更大搜索范围的情况下,SiamFC 可以在目标经历快速移动和遮挡时成功地跟踪目标。然而,没有在线的模型更新过程导致 SiamFC 对于相似的外观模型非常敏感,对于光照的变化没有特殊的处理,从而错误地跟踪到相似区域如视频流花园和视频流光饱和所示。本文提出的 TCOOT 跟踪算法在 4 个视频流中各项干扰下都能准确跟踪目标,跟踪速度保持在 46.7 frames/s,在实际应用过程中也保有很高的实时性。

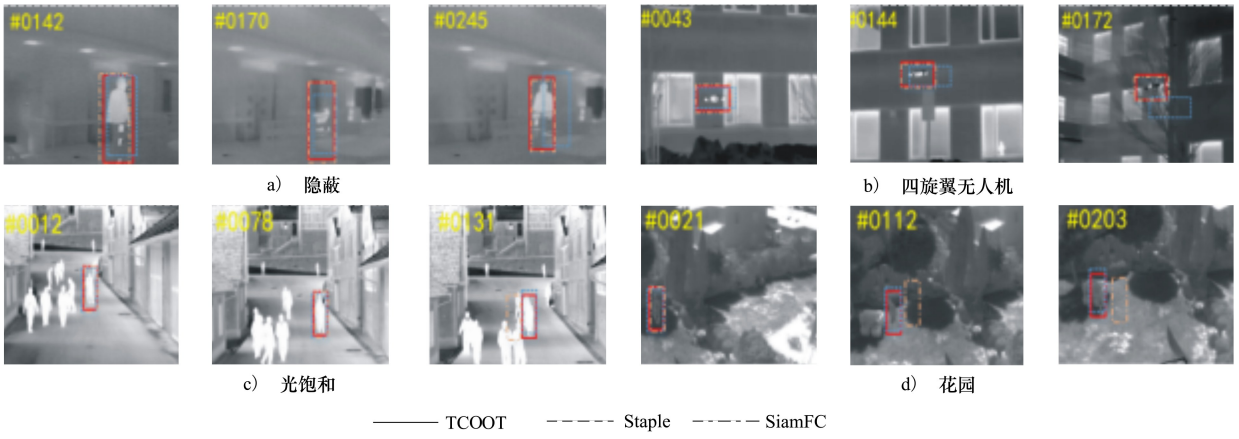


图 6　TCOOT,Staple 和 SiamFC 跟踪算法的截图

4　结　论

本文针对红外目标跟踪提出在线离线模型更新相结合的算法。该算法不仅可以在目标跟踪过程中实时更新外观模型,同时具备在离线阶段由大量图像数据训练得到的鲁棒模型。与传统的在线更新算法不同,本文提出算法并非通过增加外观模型的复杂性来避免跟踪错误,而是应用更加有效的跟踪算法(例如 Staple)来保持实时处理。本文提出利用联合响应图的方式来检测在目标跟踪过程中可能的干扰,并通过基于响应图的离散度测量方式和自适应阈值来实现在线模型更新与离线模型更新之间的切换以保证在特定情境干扰时算法的鲁棒性。本文算法在 VOT-TIR-2015 数据库上与 3 种其他算法比较,体现了算法的有效性。

参考文献:

[1] Felsberg M, Berg A, Hager G, et al. The Thermal Infrared Visual Object Tracking VOT-TIR2015 Challenge Results[C]//Proceedings of IEEE International Conference on Computer Vision Workshop, 2015: 1639-1673

[2] Hare S, Golodetz S, Saffari A, et al. Struck: Structured Output Tracking with Kernels[J]. IEEE Trans on Pattern Analysis & Machine Intelligence, 2016, 38(10): 2096-2109

[3] Danelljan M, Khan F S, Felsberg M, et al. Adaptive Color Attributes for Real-Time Visual Tracking[C]//Proceedings of conference on Computer Vision and Pattern Pecognition, 2014: 1090-1097

[4] Bertinetto L, Valmadre J, Golodetz S, et al. Staple: Complementary Learners for Real-Time Tracking[C]//Proceedings of the Conference on Computer Vision and Pattern Recognition, 2016: 1354-1378

[5] Bertinetto L, Valmadre J, Henriques J, et al. Fully Convolutional Siamese Networks for Object Tracking[C]//Proceedings of the European Conference on Computer Vision, 2016: 3376-3383

[6] Vedaldi A, Lenc K. MatConvNet-Convolutional Neural Networks for MATLAB[C]//Proceedings of the ACM International Conference on Multimedia, 2015: 689-723

Real-Time Thermal Infrared Tracking Based on Collaborative Online and Offline Method

Zhang Ximing, Wang Mingang, Cao Lin

(School of Astronautics, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: Most tracking-by-detection based trackers employ the online model update scheme based on the spatio-temporal consistency of visual cues. In presence of self-deformation, abrupt motion and heavy occlusion, these trackers suffer from different attributes and are prone to drifting. The model based on offline training, namely Siamese networks is invariant when suffering from the attributes. While the tracking speed of the offline method can be slow which is not enough for real-time tracking. In this paper, a novel collaborative tracker which decomposes the tracking task into online and offline modes is proposed. Our tracker switches between the online and offline modes automatically based on the tracker status inferred from the present failure tracking detection method which is based on the dispersal measure of the response map. The present Real-Time Thermal Infrared Collaborative Online and Offline Tracker (TCOOT) achieves state-of-the-art tracking performance while maintaining real-time speed at the same time. Experiments are carried out on the VOT-TIR-2015 benchmark dataset and our tracker achieves superior performance against Staple and Siam FC trackers by 3.3% and 3.6% on precision criterion and 3.8% and 5% on success criterion, respectively. The present method is real-time tracker as well.

Keywords: thermal infrared tracking; staple; siamese network; failure tracking detection