

基于加权重叠率的 单目标视觉跟踪评价指标

孙 巧,张胜修,张正新,曹立佳,李小锋

(火箭军工程大学控制工程系,陕西西安 710025)

摘 要: 针对真值标注的歧义性、偏差性问题和具有缩放场景的视觉跟踪应用情况,提出了一种新的视觉跟踪单目标基准评价指标. 首先,在重叠率基础上提出了加权重叠率框架;其次,提出了多区域标注方法,通过多区域标注降低标注者歧义性,在具有缩放场景的应用中,通过反演进行多区域标注,使评价更符合应用实际;再次,针对标注的偏差性,提出了多标注融合方法,提高了标注的可信度;最后,将应用于单次跟踪评价的重叠率框架推广到多次跟踪评价,利用加权结果图使评价更具解释性. 通过著名评价标准 VOT、OTB 的真值标注融合实验验证了本文标注规则的准确性;通过在具有缩放场景的视觉跟踪实验和重复实验,与其他跟踪指标的比较验证了本文指标的有效性.

关键词: 视觉跟踪;评价指标;重叠率;注释;缩放

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112 (2017)03-0753-09

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2017.03.036

A Weighted-Overlap Based Metric for Single Visual Object Tracking Evaluation

SUN Qiao, ZHANG Sheng-xiu, ZHANG Zheng-xin, CAO Li-jia, LI Xiao-feng

(Department of Control and Engineering, Rocket Force University of Engineering, Xi'an, Shaanxi 710025, China)

Abstract: Aimed at the problems of annotation of ground truth and the application of zooming, a new basic metric for visual tracking evaluation is proposed. Firstly, a weighted-overlap frame is reconstructed based on the traditional overlap. Secondly, we put forward multiple region annotation to decrease the deviation and apply in zooming. Thirdly, a multi-label fusion method is presented to improve the confidence level of the labels. Last but not least, the presented methods are expanded to repeated visual tracking evaluation, where a weighted result chart is utilized to make the evaluation more explanatory. Experimental results show that our annotation rule are more accurate than VOT and OTB, and the proposed metric is more appropriate than other metric.

Key words: visual tracking; evaluation metric; overlap; annotation; zooming

1 引言

视觉跟踪是计算机视觉的重要研究内容,在视频监控、智能人机交互、医疗诊断、成像制导等多个领域具有广阔的应用前景^[1]. 但是,由于光照、遮挡、摄像机运动等因素的影响,视觉跟踪算法仍然是一个挑战性的难题^[2],众多的文献及算法使人很难掌握视觉跟踪的发展趋势. 通过研究视觉跟踪算法的好坏,来把握视觉跟踪发展脉搏,是发展和创造视觉跟踪算法一个重要途径. 近年来,相关学者开始开展视觉跟踪算法评价的

研究,计算机视觉的国际著名期刊和顶级会议持续出现较多的视觉跟踪算法评价相关报道^[3-5]. Nawaz^[6]提出了一个以重叠率为基础的、准确率和失败率有机结合的复合评价指标,在模拟真实世界的场景序列上对视觉跟踪算法进行评价. Smeulders 等人^[7]利用基于重叠率的 F 得分和中心误差来评价视觉跟踪算法. Wu 等人^[5]建立了 100 个序列、29 个先进跟踪器的较大型视觉评价基准 (OTB, Object Tracking Benchmark),通过成功率来评价视觉跟踪,在此基础上提出了时间和空间的鲁棒性评价. 然而在序列标注上,更多的是收集已有

的、含真值标注的序列集. 由 Matej Kristan 等人组织的 VOT 挑战 (Visual Object Tracking Challenge)^[7-8], 在评价指标上采用了平均重叠率来评价视觉跟踪的准确性和利用重叠率来检测的失败率评价视觉跟踪的鲁棒性. Teng 等人^[9]将中心误差与重叠率相结合, 提出了一种新的复合评价指标.

尽管提出了多种视觉跟踪评价指标, 但总的来说, 视觉跟踪单目标评价指标可分为两类: 一类是以目标中心位置误差为基础的视觉跟踪评价体系, 该类评价简单有效, 但不适合具有尺度变化的目标, 特别是对真值标注的准确性非常敏感; 另一类是以重叠率为基础的视觉跟踪评价体系. 该类评价对真值标注的准确性要求比基于目标中心位置误差的视觉跟踪评价稍低. 并且传统的重叠率似乎对视觉跟踪评价要求过于苛刻^[10]. 然而上述两种评价都不适合具有摄像机缩放的视觉跟踪评价, 且多数视觉跟踪评价平台都没有清晰定义目标的包围盒 (真值标注).

作为视觉跟踪评价中的关键步骤^[10-11], 真值标注的歧义性^①和偏差性^②是标注的常见问题. 而在具有摄像机缩放的视觉跟踪中, 摄像机常会将需要跟踪的区域放大, 完成对区域中的某一重点子区域进行观测, 即希望在摄像机发大后, 仍然能准确跟踪目标, 这就要求视觉跟踪评价能使那些能跟踪重点区域的算法脱颖而出, 而传统的评价指标将每个像素视为相同重要性, 不能体现重点区域的重点性.

为此, 本文提出了加权重叠率. 利用多区域标注和多标注融合方法来降低视觉标注的歧义性和偏差性, 并将多区域标注应用于缩放场景视觉跟踪评价. 同时在多次重复实验中, 应用加权结果图对跟踪算法进行评价, 获得了更直观的视觉跟踪评价结果.

2 评价指标

2.1 评价指标的框架

重叠率 S 表征跟踪器结果中的目标区域 R^{RT} 与目标真值区域 R^{GT} 的重叠比率, 具体为:

$$S = \frac{|R^{RT} \cap R^{GT}|}{|R^{RT} \cup R^{GT}|} \quad (1)$$

其中 \cap 、 \cup 表示集合的交、并集操作, $|\cdot|$ 表示区域中像素的个数.

在像素分类^[10]中, 该重叠率为:

$$S = \frac{|R^{RT} \cap R^{GT}|}{|R^{RT} \cup R^{GT}|} = \frac{|TP|}{|TP| + |FN| + |FP|} \quad (2)$$

其中, GT 表征目标真值包围框, RT 表征跟踪器目标结果包围框, TP 表征 R^{RT} 和 R^{GT} 的重叠像素, FN 表示除去 TP 剩余属于 R^{RT} 的像素, FP 表示除去 TP 剩余属于 R^{GT} 的像素. 在重叠率基础上, 对目标真值区域和目标结果区域的

每个像素赋予一定的加权, 则提出的加权重叠率为:

$$S^w = \frac{|(w(T)R^{RT}) \cap (w(GT)R^{GT})|}{|(w(T)R^{RT}) \cup (w(GT)R^{GT})|} \quad (3)$$

其中, $w(T)$ 、 $w(GT)$ 分别表征 R^{RT} 和 R^{GT} 的加权.

同理, 该重叠率可为:

$$S^w = \frac{\sum_i^{|TP|} w_i(TP)}{\sum_i^{|TP|} w_i(TP) + \sum_i^{|FN|} w_i(FN) + \sum_i^{|FP|} w_i(FP)} \quad (4)$$

其中, $w_i(TP)$ 、 $w_i(FN)$ 、 $w_i(FP)$ 分别表示属于 TP 、 FN 、 FP 的第 i 个像素的加权.

2.2 评价指标的讨论

若将像素加权均等化, 即 $w_i(TP) = w_i(FN) = w_i(FP)$, 则式(4)改写为:

$$S^w = \frac{|TP|}{|TP| + |FN| + |FP|} \quad (5)$$

显然, 该加权重叠率等效为重叠率. 因此, 本文提出的加权重叠率可看成是重叠率的扩展.

如果令 $w_i(TP) = 2w_i(FN) = 2w_i(FP)$, 则式(4)简化为:

$$S^w = \frac{2|TP|}{2|TP| + |FN| + |FP|} \quad (6)$$

该加权重叠率等效为信息检索的 F 度量^[10].

由讨论可知, 通过赋予不同的加权, 加权重叠率具有不同的表现形式, 因此确定每个像素的加权是非常重要的.

2.3 评价指标的实施

2.3.1 单标注

单标注是指标注由一人一次性完成. 不同于仅在视觉图像单一区域标注的传统方法, 本文对视觉图像多个区域进行等级标注. 具体步骤如下:

(1) 多区域等级标注

即在视觉跟踪图像不同区域进行不同等级的标注. 在人工标注的实践中, 首先根据先验知识和任务需求, 确定标注的区域并赋予其权重. 权重主要根据先验知识和具体应用获得, 先验知识主要包括: 人们对目标的认识、任务需求的相关知识等, 具体应用主要是指检测的需要标注的目标类型、任务需求等. 在权重赋予过程中, 既可对标注区域每一像素赋予相同的权值 (为计算方便, 一般取整数), 也可对所标注的区域赋予区域的总权值, 进而求得各个像素的加权重. 如图 1 所示对单个行人进行标注, 首先按照传统方法对整个行人进

① 一般的, 标注选用规则框 (如矩形框), 一些狭小目标部分或一些易变的部分, 常使视觉跟踪者难以决策, 一方面标注上, 有较多的背景像素, 不标注, 又有显著的目标元素, 如何抉择让标注者“左右为难”, 即标注的歧义性.

② 由于人眼的差异及思考习惯, 会使标注出现偏差, 即标注的偏差性.

行标注(如 A_1),赋予每个像素权重为 1,然后对上半身进行标注(如 A_2),重新赋予每个像素权重为 2,这样对于显著不变的上半身赋予较高的权重,而对易变的腿部区域赋予较低的权重.也可通过类似文献[12]方法对目标进行分类,通过构造不同部分的检测器,完成自动标注,其权重仍然利用先验知识获得.

(2) 权重存贮

类似于传统标注的存贮仅存贮真值包围盒,该标注存贮权重及其权重所占包围盒.如图 1 所示,存贮为

$$[w(1), x_1, y_1, w_1, h_1, w(2), x_2, y_2, w_2, h_2],$$

其中 x_1, y_1, w_1, h_1 和 x_2, y_2, w_2, h_2 分别表示低权重和高权重包围盒的区域、 $w(1)$ 和 $w(2)$ 表示其权重.

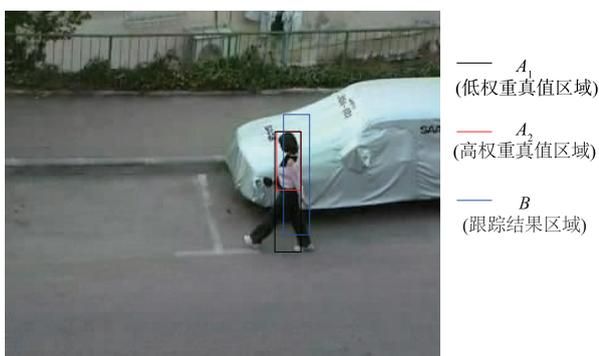


图1 单标注示意图

(3) 计算加权重叠率

通过存储的权重和区域,利用式(6)计算加权重叠率.视觉常见加权标注如图 1 所示,则式(4)可利用面积计算.

$$S^w = \frac{F(T_1) + (w_{a2} - 1)F(T_2)}{F(A_1) + (w_{a2} - 1)F(A_2) + F(B) - F(T_1)} \quad (7)$$

其中 A_1, A_2 分别表示低权重和高权重真值区域,且 $A_2 \subset A_1$,权重分别为 $w_{a1} = 1, w_{a2}$ (w_{a2} 主要根据先验知识和具体应用获得); B 表示跟踪结果区域, F 表示区域框下的面积; $T_1 = A_1 \cap B, T_2 = A_2 \cap B$.

若无特定的应用环境,取 $w_{a2} = 2$,则有:

$$S^w = \frac{F(T_1) + F(T_2)}{F(A_1) + F(A_2) + F(B) - F(T_1)} \quad (8)$$

在具有缩放的视觉跟踪评价中,利用缩放后需要跟踪的区域反演推出缩放前重点跟踪区域,赋予较高的权值,一般地,该区域的权重可由该区域的重要度函数求得:

$$w_{R1} \propto \text{importance}(R1) \quad (9)$$

其中 importance 表示目标的重要度函数,由先验知识获得.依据上面步骤完成具有缩放的视觉跟踪评价,如图 2 所示.

2.3.2 多标注

多标注来源于已公开的真值集、多个标注者标注

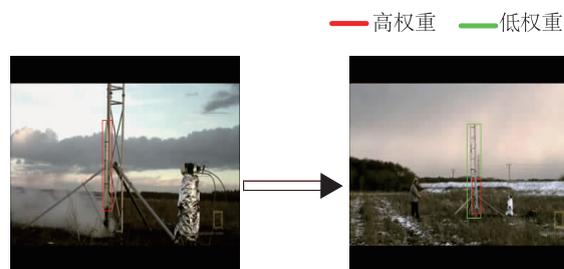


图2 缩放标注示意图

的真值集、单个标注者多次标注(标注可使用本文方法,也可用其他方法)形成的真值集以及它们的任意组合.对每个标注源,赋予一定的权重.权重可均匀给定,也可根据来源可靠性分级给定.例如:将著名实验室、优秀标注者的真值标注集的权重赋予较高的值,其它赋予较低的值.也可以仿照 Pang 等人^[3]的算法,根据其在期刊使用的频数,利用 Page-Rank 等排名算法赋予权重.

对于标注源,首先对其图像进行数值化(未标注的背景区域定位 0,传统方法标注的目标区域定位 1,多区域标注方法目标区域定位为各自等级),由于在视觉标注中标注的范围是已知的,因此图像数值标注矩阵可简化生成一个较低维的矩阵.然后将其进行加权和,获得一个真值加权矩阵,最后在视觉跟踪评价中,利用式(4)进行求解计算.图 3 显示了多标注加权过程,其中子图(a)、(b)为传统标注源,子图(c)为多区域等级标注源,子图(d)为真值加权矩阵.

为便于存贮计算,提出了简化算法,如图 4 所示.

(1) 阈值化

为避免标注带来的随机性误差,对生成的矩阵进行阈值化(如图 4(a)所示,阈值取 3),阈值可以根据多标注的个数来确定,可取总个数的 10% ~ 30%,也可以对加权后的像素值利用机器学习算法分类得到.设定阈值 ε_0 ,有

$$\begin{cases} x^a = 0, & x < \varepsilon_0 \\ x^a = x, & x \geq \varepsilon_0 \end{cases} \quad (10)$$

(2) 区间量化

为便于计算,对生成的矩阵进行区间量化(如图 4(b)所示,量化数 n 取 2,量化阈值为 6).量化数和量化区间既可人工直接给定,也可利用监督或半监督学习得到.式(11)显示了区间量化公式、 $\varepsilon_1 \cdots \varepsilon_{n-1}$ 为阈值, $\beta_1 \cdots \beta_n$ 为量化后赋予的值.

$$\begin{cases} x^b = \beta_1, & x^a < \varepsilon_1 \\ x^b = \beta_2, & \varepsilon_1 \leq x^a < \varepsilon_2 \\ \dots & \dots \\ x^b = \beta_n, & x^a > \varepsilon_{n-1} \end{cases} \quad (11)$$

(3) 规则化

为了便于存贮和直观表示目标,对生成的矩阵进行规则化,使其具有一般的表现形式(如图4(c)所示,对生成的矩阵进行了矩形化,红色、浅蓝色矩形框分别表示高权重、低权重的矩阵,红色数字,表示矩形化过程中修改的像素值).规则化,既可根据上述区间量化人工标定,也可利用监督学习和无监督学习自动标定.

(4) 存贮与计算

类似于单标注,对规则化的矩阵进行存贮,计算相应的重叠率.如图4(d)所示,标注存贮为 $[1,1,1,8,8,2,3,3,4,4]$,跟踪结果(黑色矩形框)表示为 $[0,4,6,6]$,依据式(10)则可计算出对应的加权重叠率为37.4%.

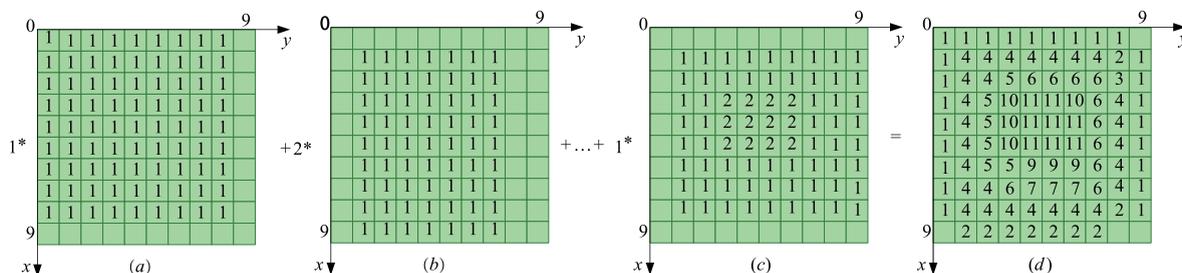


图3 多标注加权示意图

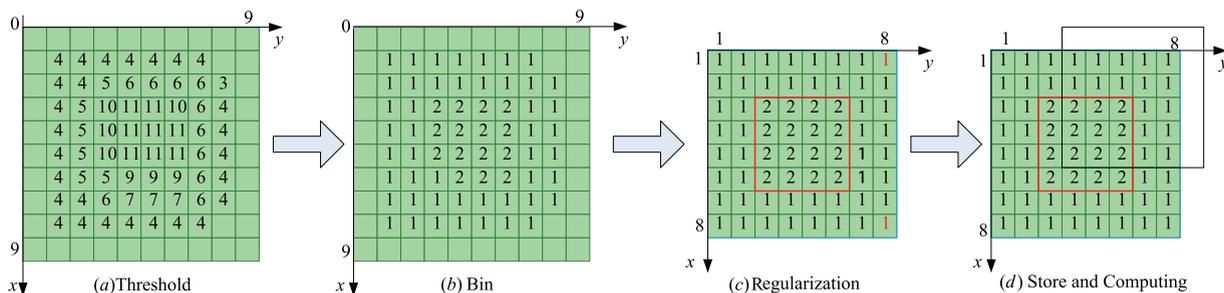


图4 多标注简化策略

2.4 评价指标的推广

在多次重复实验中,传统的评价方法需要对每次实验进行单独评价,然后根据评价结果获得一个评价统计量.利用加权重叠率可直接获得总体评价.类似多标注方法,将单次跟踪结果视为一个标注(权重均为1),按照2.3.2节所示步骤进行处理.与传统方法相比,本文方法直接获得跟踪结果统计信息,然后再进行评价,因此评价更具直观性和可解释性.

若 B_1, B_2 分别表示目标跟踪结果低权重和高权重区域,且 $B_2 \subset B_1$,权重为 $w_{b1} = 1, w_{b2}$ (由区间量化可得), A 表示目标真值区域, $T_3 = A \cap B_1, T_4 = A \cap B_2$,则式(4)为

$$S^w = \frac{F(T_3) + (w_{b2} - 1)F(T_4)}{F(A) + F(B_1) + (w_{b2} - 1)F(B_2) - F(T_3)} \quad (12)$$

从本质上讲,2.3节确定为加权真值图,2.4节确定的为加权结果图,而视觉跟踪评价就是评价两者之间的相关性.若 $A_2 \subset A_1, B_2 \subset B_1, T_5 = A_1 \cap B_1, T_6 = A_1 \cap B_2, T_7 = A_2 \cap B_1, T_8 = A_2 \cap B_2$,则式(4)为

$$S^w = \frac{T_i}{T_u}$$

$$T_i = F(T_5) + (w_{a2} - 1)F(T_6) + (w_{b2} - 1)F(T_7) + (w_{a2} - 1)(w_{b2} - 1)F(T_8)$$

$$T_u = F(A_1) + F(B_1) + (w_{a2} - 1)F(A_2) - F(T_5) + (w_{b2} - 1)F(B_2) + (w_{a2} - 1)(w_{b2} - 1)F(T_8) \quad (13)$$

若无特定的应用要求,取 $w_{a2} = 2, w_{b2} = 2$ 时,有

$$S^w = \frac{F(T_5) + F(T_6) + F(T_7) + F(T_8)}{\sum_{k=1}^2 [F(A_k) + F(B_k)] - F(T_5) + F(T_8)} \quad (14)$$

2.5 评价指标的分析

2.5.1 标注不准确的影响

图5、图6和图7分别显示了在标注出现偏移、尺度偏大、尺度偏小的情况,对 GT_m 所产生的影响(GT_m 表示人工标注获得的真值, GT 表示理论真值,高权重像素权重是其它像素权重的两倍).左图为利用加权标注的 GT_m 和 GT ,右图为普通标注的 GT_m 与 GT .

通过计算可得,图5~图7左图 GT_m 与 GT 的重叠比率分别为75.5%、86%、44.9%,而右图 GT_m 与 GT 的

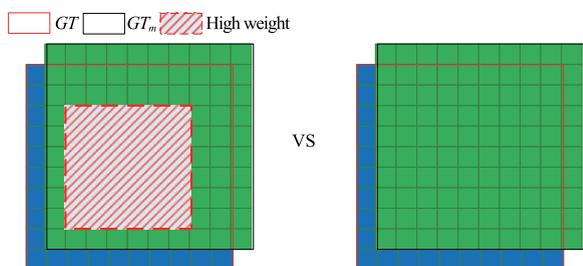


图5 偏移不准确的影响

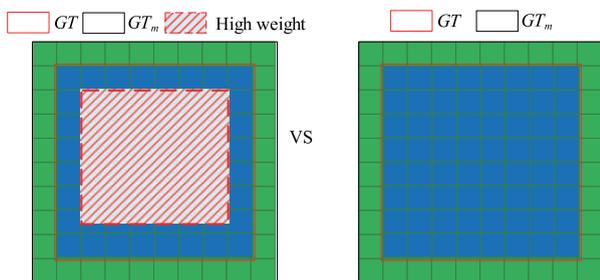


图6 尺度偏大的影响

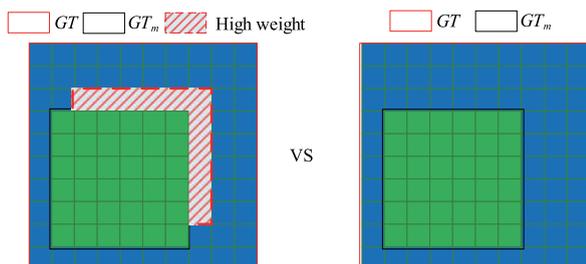


图7 尺度偏小的影响

重叠比率为 68.1%、81%、36%。这说明当人工真值与理论真值出现相同偏差时,利用本文获得标注比传统标注更能靠近理论真值,以本文标注的真值为基础的视觉跟踪算法评价的可信度将更高。

2.5.2 克服传统重叠率的局限

图 8 显示了对传统重叠率局限克服^[9]的示意图。左图显示了 $RT \subset GT$, 3 个跟踪器的传统重叠率均为 12%, 不能表征跟踪器性能的差异, 而加权重叠率则为 14.7%、13.2%、10.3% (由上至下), 更符合评价实际。同理, 中图显示了 $GT \subset RT$, 传统重叠率和 2.3 节加权真值图不能评价两个跟踪器性能, 但是利用 2.4 节提出加权真值图和加权结果图能够很好地解决这一问题, 如右图所示。从图 8 可以看出, 当传统重叠率不能评价跟踪器时 (当 $RT \subset GT$ 或 $GT \subset RT$), 利用加权重叠率则能很好评价跟踪器。

3 实验

3.1 真值标注融合对比实验

本节利用著名评价标准 VOT^[7]、OTB^[5] 真值标注集

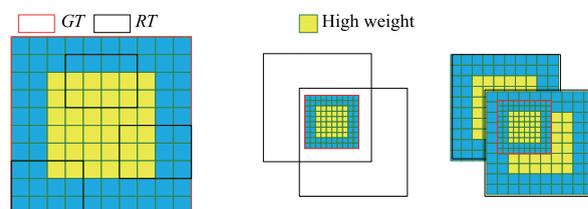


图8 克服传统重叠率的局限

的融合结果 (量化数取 2) 来展示多标注融合有效性 (如图 9 所示)。

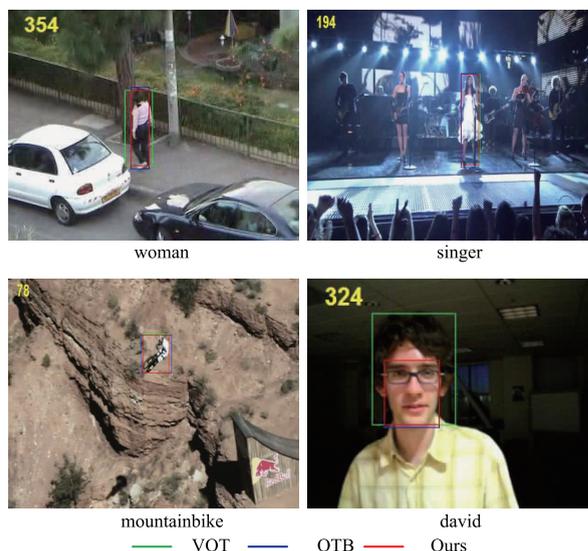


图9 真值标注对比图

由图 9 可知, 尽管 VOT、OTB 广泛应用于各种先进的跟踪器评价中, 但其标注仍然具有差异性。这种差异反映了真值的不真值性, 一方面来源于标注规则的选取, 即真值标注的歧义性, 如在“david”序列, VOT 将头发标注, 而 OTB 则仅标注眼睛到下颏位置。另一方面也来源于真值标注的偏差性, 如“woman”、“singer”、“mountainbike”两者出现的稍许偏差。利用多标注融合可提高标注的准确性, 增加标注的可信度, 如“woman”、“singer”、“mountainbike”关键位置进行标注、更贴合实际目标位置。而为了更好地增加标注的可信性, 需要收集更多的标注集, 这也是未来的工作之一。

3.2 具有缩放场景的视觉跟踪实验

为了展示提出的评价指标的有效性, 利用 ACAT^[13]、ASLA^[14]、KCF^[15]、DFT^[16]、OSPT^[17]、STC^[18] 等 6 个算法在具有缩放场景的序列图像“zooming” (来源于优酷视频) 对卡车进行跟踪, 同时用中心误差^[10]、传统的重叠率^[5]、基于中心误差与重叠率的复合指标^[9]与加权重叠率 (标注卡车时, 车头为重要区域, 权重取 2, 其他卡车部分为非重要区域, 权重取 1) 进行跟踪算法的评价, 完成评价指标的比较。图 10 显示了 6 种

算法在“zooming”序列中的跟踪结果.



图10 在zooming序列跟踪算法的结果图

由图 10 可知,在无缩放场景下,6 种算法均能较好跟踪目标(帧 28),然而在缩放比例发生变化时,所有跟踪器都出现不同程度性能下降的问题(如帧 5 帧显示)跟踪效果.变尺度 ASLA 性能始终较好,而变尺度 STC 却出现了下降,而固定尺度 ACAT、OSPT 却在 819 帧表现较好,这说明在缩放场景中,合适的自适应尺度变化将提升视觉跟踪效果,但是在无任何先验知识的情况,固定尺度反而展现了更强的鲁棒性.表 1~表 2 显示了 6 种算法在“zooming”序列中代表帧的跟踪结果.

表 1 Zooming 序列 437 帧的跟踪算法评价表

评价指标	ACAT	ASLA	DFT	KCF	OSPT	STC
中心误差	23.77	10.30	3.606	24.59	20.62	37.01
重叠率	0.032	0.399	0.032	0.032	0.032	0.147
复合指标	0.025	0.360	0.031	0.025	0.026	0.089
加权重叠率	0.043	0.430	0.028	0.049	0.042	0.234

表 2 Zooming 序列 819 帧的跟踪算法评价表

评价指标	ACAT	ASLA	DFT	KCF	OSPT	STC
中心误差	8.062	4.472	9.849	5.831	8.062	8.062
重叠率	0.233	0.194	0.092	0.233	0.233	0.067
复合指标	0.113	0.149	0.041	0.144	0.113	0.051
加权重叠率	0.321	0.180	0.165	0.298	0.321	0.090

从表 1 到表 2 可知,基于中心误差不适合目标尺度

变化的跟踪评价(表 1 中以中心误差来看,DFT 跟踪效果最好,但结合图 10 可知,DFT 尺度明显偏小,跟踪效果不如 ASLA).复合指标始终比重叠率低(表 1 和表 2 所示),不利于改善重叠率对视觉跟踪要求过于苛刻的弊病.传统重叠率容易陷入到当 $RT \subset GT$ 时,不能评价跟踪算法的局限,而加权重叠率,能够克服传统重叠率的局限,也能表征重点区域的重要性,更适合评价具有缩放场景的视觉跟踪(表 1 中,ACAT、DFT、KCF、OSPT 重叠率相同.但中心误差、图 10 的跟踪效果和加权重叠率明显不同,从中心误差来看,DFT 跟踪较好,而从图 10 的跟踪效果来看,KCF 更靠近重点区域,加权重叠率表征了这一结果).

3.3 跟踪器多次重复评价

为了展示加权重叠率所增加的解释性,采用 VOT 提供的 CT、HT、LGT、MIL、ORIA、PJS-S 等 6 种跟踪算法在序列“bicycle”第 85 帧的运行 15 次的结果^[7],利用中心误差^[19]、传统的重叠率^[5]与加权重叠率进行评价,完成评价指标的比较.加权重叠率参数设置如下:量化数均取 $n=2$,加权重叠率 1 阈值 $\varepsilon_0=2, \varepsilon_1=10, w_{a2}=2$;加权重叠率 2 阈值 $\varepsilon_0=2, \varepsilon_1=10, w_{a2}=3$;加权重叠率 3 阈值 $\varepsilon_0=1, \varepsilon_1=10, w_{a2}=2$;加权重叠率 4 阈值 $\varepsilon_0=3, \varepsilon_1=10, w_{a2}=2$.

图 11 展示了 6 种算法在序列“bicycle”第 85 帧跟踪结果图(前 15 帧为算法结果,第 16 帧 VOT 提供的标注真值).图 12 显示了利用本文方法获得加权结果图.对比图 11、图 12 可知,传统方法在单帧中,可以较为清晰地分辨出各种算法的性能,然而在多帧图像中却不能很直观反应单个跟踪算法的性能,而加权结果图可直观看到 15 次算法性能.如 HT,高权值结果框与低权值结果框紧密贴合一块,而 CT 相距较远,以此可推出在 HT 的随机性较小、CT 的随机性较大. PJS-S 高权值、低权值结果框均距真值框较近,特别是高权值框离真值框在 6 个算法中最近,以此可推出在 PJS-S 的跟踪效果最好.这与表 3 得到的结果一致.

表 3 Bicycle 序列 85 帧的跟踪算法评价表

评价指标	CT	HT	LGT	MIL	ORIA	PJS-S
中心误差	9.270	2.450	6.164	4.494	8.729	2.049
重叠率	0.454	0.488	0.529	0.336	0.629	0.734
加权重叠率 1	0.740	0.614	0.665	0.508	0.700	0.922
加权重叠率 2	0.796	0.673	0.708	0.604	0.729	0.943
加权重叠率 3	0.740	0.608	0.640	0.508	0.640	0.819
加权重叠率 4	0.747	0.614	0.695	0.508	0.769	0.979

表 3 显示了 6 种算法的评价表,从评价指标可知, PJS-S 跟踪效果最好.对比重叠率与加权重叠率可知,6 种跟踪算法的加权重叠率均比重叠率高,特别是在 PJS-S

能够很好跟踪目标情况下,重叠率却仅为 73.4%,而加权重叠率 1 为 92.2%。利用加权重叠率进行算法评价,更能与人的直观解释相结合,以此为基础的跟踪成功率的判断,将更有说服力,从而改善了传统重叠率在跟踪成功率的阈值过高的缺陷。对比加权重叠率 1 和 2 可以看出,提

高重点区域的权重,跟踪较好的算法重叠率将继续提高。对比加权重叠率 1、3 和 4 可知,不同的阈值,对应不同的加权重叠率,不同跟踪算法对其阈值的敏感度也不同,MIL 具有最高的集聚性,即算法稳定,PJS-S 虽然跟踪较好,仍然具有一定发散性。

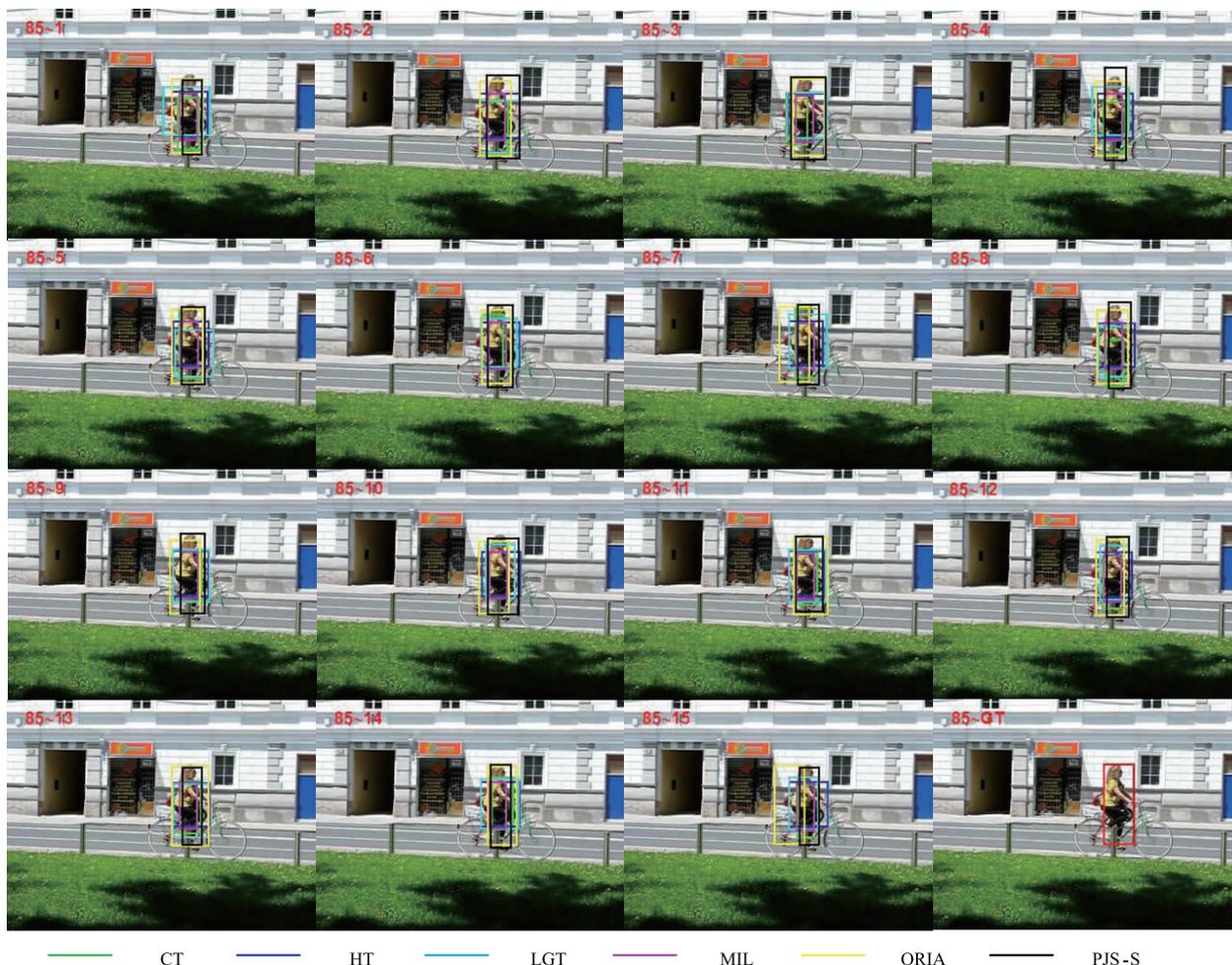


图11 序列bicycle第85帧跟踪算法的结果图

表 4 显示了 6 种算法的跟踪结果分别基于 VOT、OTB 的标注下评价结果,其中,重叠率 1、2 分别表示基于 VOT、OTB 真值标注的重叠率评价结果,加权重叠率 1 则是利用本文方法(式 14)获得综合性评价结果。重叠率 1 和重叠率 2 所采用的跟踪结果完全一样,评价方法完全一样,但是标注出现微小偏差,也会对跟踪结果评价产生较大影响。而利用加权重叠率则能给出唯一的评价结果。

表 4 基于不同标注的 singer 序列 85 帧的跟踪算法评价表

评价指标	CT	HT	LGT	MIL	ORIA	PJS-S
重叠率 1	0.601	0.332	0.305	0.645	0.717	0.551
重叠率 2	0.595	0.356	0.323	0.659	0.650	0.494
加权重叠率	0.750	0.506	0.441	0.817	0.815	0.728

4 小结

本文以重叠率为基础提出了加权重叠率,通过加权重叠率来评价视觉跟踪算法,对真值标注的歧义性、偏差性和摄像机缩放均具有较好的效果,并得到了如下见解:①缩放场景中,一些在传统场景表现优异的跟踪器效能明显下降,需要专门针对缩放进行研究;②合适的自适应尺度将提升跟踪效果,但在无任何先验知识的情况下,固定尺度反而展现了更强的鲁棒性;③低权重的像素对于跟踪也很重要。

然而,本文在单帧讨论了基于规则框的基准视觉跟踪评价指标,而完整视觉跟踪评价需要在多个序列图像集进行评价,也不适合形状变化明显的非刚体目



图12 视觉跟踪算法的加权结果图

标跟踪算法评价。因此,未来工作和研究方向如下:①以本文的基准评价指标为基础,构建视觉评价指标体系,并选用标准的数据库和基准的跟踪器,开发视觉跟踪评价平台;②由规则框扩展到基于轮廓等不规则框的视觉跟踪算法评价;③在多标注融合中,进一步综合利用图像表观信息,提出更加智能的评价方法。

参考文献

- [1] 李康,何发智,潘一腾,等. 基于簇相似的多分类器目标跟踪算法[J]. 电子学报,2016,44(4):821-825.
LI Kang, HE Fa-zhi, PAN Yi-teng, et al. Multi-classifier object tracking based on cluster similarity[J]. Acta Electronica Sinica,2016,44(4):821-825. (in Chinese)
- [2] 宋涛,李鸥,刘广怡. 基于置信区域内多级动态层表达的类序蒙特卡洛视觉跟踪方法[J]. 电子学报,2016,44(6):1355-1361.
SONG Tao, LI Ou, LIU Gang-yi. Quasi-sequential Monte Carlo visual tracking based on multilevel dynamic layer representations in confidence region[J]. Acta Electronica Sinica,2016,44(06):1355-1361. (in Chinese)
- [3] PANG Y, LING H. Finding the best from the second best-inhibiting subjective bias in evaluation of visual tracking algorithms[A]. Proceedings of the 2013 IEEE International Conference on Computer Vision[C]. Sydney, NSW, Australia:IEEE Press,2013. 2784-2791.
- [4] SMEULDE AWM, CHU DM, CUCCHIARA R, et al. Visual tracking: an experimental survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(7):1442-1468.
- [5] WU Y, LIM J, YANG M. Object tracking benchmark[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9):1834-1848.
- [6] NAWAZ T, CAVALLARO A. A protocol for evaluating video trackers under real-world conditions [J]. IEEE Transactions on Image Processing, 2013, 22(4):1354-1361.
- [7] KRISTAN M, PFLUGFELDER R, LEONARDIS A, et al. The visual object tracking VOT2013 challenge results[A]. Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops [C]. Sydney, NSW, Australia:IEEE Press,2013. 98-111.
- [8] KRISTAN M, MATAS J, LEONARDIS A, et al. A novel performance evaluation methodology for single-target trackers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(11):2137-2155.
- [9] TENG H, YU L, LIU H, et al. A novel performance evaluation method for visual tracking methods[A]. Proceedings of the 2015 Chinese Intelligent Automation Conference [C]. Fuzhou China:Springer,2015. 313-324.
- [10] CEHOVIN L, LEONARDIS A, KRISTAN M. Visual object tracking performance measures revisited [J]. IEEE Transactions on Image Processing, 2015, 25(3):1261-1274.
- [11] KASTURI R, GOLDFOG D, SOUNDARARAJAN P, et al. Framework for performance evaluation of face, text, and vehicle detection and tracking in video: Data, metrics, and protocol[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2):319-336.
- [12] 王国华,刘琼,庄家俊. 基于局部特征的车载红外行人检测方法研究[J]. 电子学报, 2015, 43(7):1444-1448.
WANG Guo-hua, LIU Qiong, ZHU Jia-jun. Method research on vehicular infrared pedestrian detection based on local features[J]. Acta Electronica Sinica, 2015, 43(7):1444-1448. (in Chinese)
- [13] DANELLJAN M, SHAHBAZ KHAN F, FELSBERG M, et al. Adaptive color attributes for real-time visual tracking [A]. IEEE Conference on Computer Vision and Pattern Recognition [C]. USA:IEEE Press,2014. 1090-1097.
- [14] JIA X, LU H, YANG M. Visual tracking via adaptive structural local sparse appearance model[A]. Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition [C]. Providence, RI, USA: IEEE Press,2012. 1822-1829.
- [15] HENRIQUES JF, CASEIRO R, MARTINS P, et al. High-speed tracking with kernelized correlation filters[J]. IEEE Transactions on Pattern Analysis and Machine Intelli-

gence, 2015, 37(3): 583 - 596.

- [16] SEVILLA-LARA L, LEARNED-MILLER E. Distribution fields for tracking [A]. Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition [C]. Providence, RI, USA: IEEE Press, 2012. 1910 - 1917.
- [17] WANG D, LU H, YANG M. Online object tracking with sparse prototypes [J]. IEEE Transactions on Image Processing, 2013, 22(1): 314 - 325.
- [18] ZHANG K, ZHANG L, LIU Q, et al. Fast visual tracking via dense spatio-temporal context learning [A]. Proceedings of the 2014 European Conference on Computer Vision [C]. Florence, Italy, 2014. 127 - 141.
- [19] 余旺盛, 田孝华, 侯志强, 等. 基于局部分块学习的在线视觉跟踪 [J]. 电子学报, 2015, 43(1): 74 - 78.
YU Wang-sheng, TIAN Xiao-hua, HOU Zhi-qiang, et al. Online visual tracking based on local patch learning [J]. Acta Electronica Sinica, 2015, 43(1): 74 - 78. (in Chinese)

作者简介



孙 巧(通信作者) 男, 1988 年 04 月出生于湖北荆门. 2012 年于第二炮兵工程大学获得硕士学位, 现为火箭军工程大学博士研究生, 主要从事机器视觉及智能控制方面的研究.
E-mail: seq1211@126.com



张胜修 男, 1963 年 6 月出生于陕西西安. 博士, 教授, 博士生导师, 主要从事导航、制导与控制方面的研究.
E-mail: zsx1963@aliyun.com.cn