

基于双矩形框标定的改进 TLD 算法

张伟伟 汤光明 孙怡峰

(信息工程大学 郑州 450001)

摘要 鉴于基于单个矩形框标定的 Tracking-Learning-Detection (TLD) 算法无法兼顾跟踪目标的“重点性”和“完整性”,提出了一种基于双矩形框标定的改进算法。在标定整个目标区域的矩形框的基础上,算法在目标变化相对稳定的区域标定另一个矩形框,以指示跟踪的重点区域。在提取跟踪点时,采用分配权重的方法使重点区域产生更多的跟踪点,从而提高算法对局部变化的适应能力。实验表明,改进后的算法在跟踪局部保持稳定而其余部分有所变化的目标上有较高的性能提升;而对于不存在稳定局部区域的目标,跟踪效果没有明显改善。

关键词 TLD, 区域标定, 物体跟踪, 矩形框

中图分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.09.064

Improved TLD Algorithm Based on Region Marking of Double Bounding Boxes

ZHANG Wei-wei TANG Guang-ming SUN Yi-feng

(Information Engineering University, Zhengzhou 450001, China)

Abstract Considering that the Tracking-Learning-Detection (TLD) framework can hardly balance the tracking of the whole and the interest, by using region-marking method of single bounding box, this paper proposed an improved algorithm introducing a region-marking method of double bounding boxes. With one bounding box labeling the whole object, the algorithm draws another bounding box over the stable area of image to indicate the interested part. While extracting the trace points, the weighting approach is adopted to produce more points in interested area, which improves TLD's adaptability to local variance. Experimental results show that the improved algorithm has good performance in tracking the object when a fixed part keeps stable but the rest varies. For the object not containing the stable local area, the effects are not so obvious.

Keywords TLD, Region marking, Object tracking, Bounding box

1 引言

物体跟踪是机器视觉研究的一个重要课题,其目的是根据首帧初始化的目标位置,短期或长期估计目标在视频序列中的位置及运动参数。在数十年的研究中,国内外研究者针对不同需要提出了诸多跟踪算法^[1],如 IVT、VRT、FragT、BoostT、SemiT、BeSemiT、MILT、VTD 以及 TLD 等等。但由于摄像头移动、背景变化、目标形态不稳定、运动模糊、场景多样等因素影响,这些算法都有一定的局限性,它们针对某些问题和场景有效,却不能解决所有问题。比如, FragT 方法可以处理遮挡,但不能处理平面内的旋转; IVT 方法不适宜长期跟踪,也不适合跟踪位置变化较大的目标; VRT 不能应对尺寸变化问题,因此在背景复杂的场景里表现不佳。

随着智能化技术的推进,涌现出了更多富有挑战性的基于物体跟踪的现实应用,尤其是在跟踪目标先验知识未知情况下进行长期跟踪,对跟踪的准确度和适应性提出了更高的要求。比如,在智能监控应用中,为了对潜在不当行为进行预防,往往需要对临时目标进行长期跟踪,为进一步进行行为分

析做准备。这就为物体跟踪引入了两个必须解决的难题^[2]:

- 1) 先验知识缺失。由于跟踪目标是潜在的甚至随机的,跟踪前无法获取其先验知识,这要求算法必须具备一定的泛化性。
- 2) 获得的视频帧情况复杂。跟踪的长期性使得获得的视频帧变长,进而使得视频中跟踪目标出现遮挡、消失、再出现等情况的概率增大,这要求算法具备自恢复能力和在线学习能力。在诸如交通模式分析、人机体感交互、视频标注等等应用中,这样的问题同样存在。

对此, TLD 算法具有一定的优势^[2-4]。它是一个基于追踪、检测 and 学习的长期跟踪未知物体的算法,将检测器和跟踪器结合在一起,并且能够随时对跟踪结果进行评估和学习,从而实现更有效的长期跟踪。同时,这个算法框架的鲁棒性很强,具备很强的学习恢复能力。不过 TLD 还有很多需要改进和增强的地方,如不支持多物体的跟踪、对背景凌乱的状况处理不佳、应对物体被遮挡的能力不足等。另外, TLD 中机器对跟踪目标区域的输入,是靠初始帧对其进行单个矩形框标定完成的^[2-5],但由于目标物体形状不规则,难免会出现框内包含部分背景信息的情况,使机器对目标产生理解上的错

到稿日期:2013-10-08 返修日期:2014-02-19 本文受河南省科技攻关基金(122102210047)资助。

张伟伟(1989-),男,硕士生,主要研究方向为人工智能, E-mail: chitongwei@126.com; 汤光明(1963-),女,博士生导师,主要研究方向为信息隐藏和体系对抗; 孙怡峰(1976-),男,博士,主要研究方向为信息隐藏和机器视觉。

误,这会之后的长期跟踪带来困难。为了解决这样的问题,本文提出了基于双矩形框标定的算法改进策略。

2 TLD 算法原理与分析

2.1 原理介绍

在 TLD^[2]算法中,追踪、学习和检测是算法的 3 个核心,也是其名字的由来。与传统跟踪算法相比,该算法的特色在于将跟踪算法和检测算法相结合,以解决被跟踪目标在被跟踪过程中发生的形变、部分遮挡等问题。同时,通过一种改进的在线学习机制^[4]不断更新跟踪模块的跟踪点和检测模块的目标模型,使得跟踪更加稳定、鲁棒、可靠。

对于长时间跟踪而言,需要解决的一个关键问题是:当目标重新出现在相机视野中时,系统应该能重新检测到它,并开始重新跟踪。但是,长期跟踪过程中,被跟踪目标将不可避免地发生形状变化、光照变化、遮挡等情况。在传统的跟踪算法中,首帧需先调用检测模块,检测到被跟踪目标之后开始进入跟踪模块,而此后,检测模块就不会介入到跟踪过程中。该方法的致命缺陷在于,当被跟踪目标存在形状变化或遮挡时,跟踪就很容易失败;因此,对于长期跟踪,或者被跟踪目标存在较大形状变化情况下的跟踪,一些研究者采用检测的方法来代替跟踪^[2-4]。该方法虽然在某些情况下可以改进跟踪效果,但它需要一个离线的学习过程。即在检测之前,需要挑选大量被跟踪目标的样本来进行学习和训练,并且,这些训练样本要涵盖被跟踪目标可能发生的各种形变和各种尺度变化、姿态变化和光照变化等情况。利用检测的方法来达到长时间跟踪的目的,对于训练样本的选择至关重要,否则,跟踪的鲁棒性就难以保证。

因为单纯的跟踪或者单纯的检测算法都无法在长时间跟踪过程中达到理想的效果,所以,TLD 方法考虑将两者结合,并加入一种改进的在线学习机制^[4],从而使得整体的目标跟踪更加稳定、有效。

简单来说,TLD 算法由 4 部分组成^[2-4]:跟踪模块、检测模块、学习模块和综合模块,如图 1 所示。

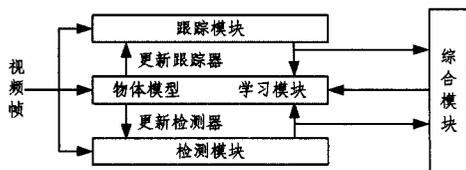


图 1 TLD 算法结构图

1)跟踪模块采用重叠块跟踪策略,单块跟踪采用 Lucas-Kanade(LK)光流法^[6,7]。开始时从标定的矩形框内取得若干跟踪点,然后通过 LK 算法追踪到这些点在两帧之间的变化情况,并将其反映到一个稀疏的运动矢量场中。基于该运动场,物体的运动状况就能够得到较好的估计。

2)检测模块^[8]由方差检测器、随机树和最近邻分类器级联构成,只有依次通过这 3 种分类器的检测区域才被认定含有检测目标。

3)综合模块将检测模块和跟踪模块得到的目标框进行综合。确定目标的最终位置时,优先考虑跟踪模块所得的结果,即如果跟踪到的图像与最初的目标图像相似度大于某阈值,就接受该跟踪结果;否则,将从检测器的结果中选出与最初目标相似度最大的图像作为跟踪结果,并更新跟踪器的最初目

标模型,用当前选用的跟踪结果替换原有的目标模型。

4)学习模块的作用是在在第一帧对目标模型进行初始化,产生正负样本并训练分类器,并且在后续的运行过程中,利用 PN 学习^[4]来对目标模型进行不断更新。

2.2 算法分析

对于长期跟踪算法来说,如果机器对跟踪目标的理解本身就存在错误,在之后的跟踪过程中错误将会不断积累,导致跟踪出现问题。TLD 算法中^[2],作者在首帧用一个矩形框对跟踪目标区域进行标定,机器将框内信息当作目标信息,提取跟踪点和特征,在后续帧中进行跟踪和检测。

但由于矩形框无法自适应不同物体的不规则形状,矩形框过大会涵盖较多背景信息,使机器误认为这些背景信息是跟踪目标的一部分,导致理解上的差错;矩形框过小可能只覆盖了目标局部,机器误把局部信息当成完整的跟踪对象,致使跟踪目标丧失了完整性,若此部分在后续帧中被完全遮挡,将导致跟踪失败。

另外,算法中矩形框的位置和尺寸是观测者自己定义的,用单个矩形框进行标定很难与跟踪目标完全契合,这些都会导致对目标物体框定不精确,进而对跟踪精度产生影响。并且,TLD 算法对矩形框位置尺寸的设定颇为敏感,框大小和位置发生变化时,可能就会使跟踪由成功变为失败。图 2 给出了 TLD 算法跟踪失败的部分视频帧,在(d)和(e)中均出现了丢帧现象。

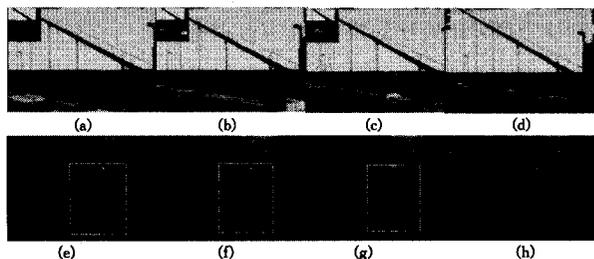


图 2 TLD 算法部分丢帧实例

3 双矩形框标定方法

目标区域标定技术旨在用坐标信息向机器解释目标所处图像和视频帧中的位置,常用于训练样本的注释^[9]和物体跟踪的初始化^[10]。

在进行物体类识别时,搜集的正样本中除含目标物体外,往往还存在大量与识别无关的背景信息,因此通常要对样本中目标位置进行注释。学习时,机器按指定位置找到物体区域,提取相应特征并送往分类器进行训练;在进行物体跟踪之前,需先明确跟踪目标的具体位置,通常在首帧对跟踪目标进行区域标定,跟踪开始后,机器将区域内信息作为跟踪对象展开跟踪。

在进行区域标定时^[11],常用的几何形状有矩形、椭圆、不规则图形等等,如图 3 所示。在应用中,矩形框最为常见,相对其它形状而言,它具有如下优势:(1)通用性强,自然场景下大部分实物轮廓都可以用矩形近似描述,如人体、杯子、文具、电器等等;(2)区域位置可直接用直角坐标系中框的左上角和右下角坐标表示,便于运算;(3)由于图片和视频帧本身为矩形,矩形框区域标定与滑动窗口、金字塔多尺度模型等方法存在更高的契合度。同时,由于矩形框无法自适应不同物体的

不同形状,其尺寸和位置又由观测者自己定义,很难恰恰框定目标本身,偏大偏小都会导致框定不精确,进而影响跟踪精度^[2,7]。

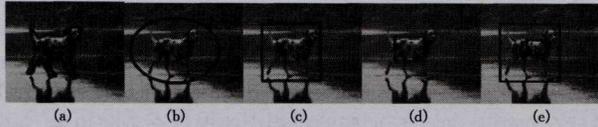


图3 目标区域标定方法

为了弥补这种缺陷,本文引入双矩形框区域标定技术,在原先标定整个物体区域的矩形框(记为BB1)基础上,再引入一个矩形框(记为BB2)以提高跟踪精度,其具体功能表现在两个方面。

1) 过滤多余背景信息

BB1 外接目标物体,涵盖目标全部信息的同时,也会或多或少地涵盖部分背景信息。目标形态越不规则,BB1 尺寸越大,包含背景信息就越多。跟踪时,这些与目标无关的背景信息可能出现剧烈变化,机器误将其当作目标特征而产生跟踪错误。此时,内切目标物体绘制尽量大的不含背景信息的矩形框 BB2,目的在于把背景信息从 BB1 中滤除掉,从而提高框定区域目标特征的稳定性。

2) 提取稳定目标信息

有的物体在被跟踪时,局部形态保持稳定,而其余部分持续变化。比如人或动物在行走移动过程中,上体一般不发生变化,而下身形态却变化剧烈。另外,在一些场景中,虽然不能精确定位发生遮挡的位置,却可以根据场景信息分析出哪个部分更易发生遮挡,即分析出哪个部分更不稳定。比如跟踪在教室行走的学生,由于受桌椅影响,其下肢被遮挡的概率比躯干和头部要高,这时为提高跟踪的准确度,应该倾向于对上体的跟踪。但如果仅把 BB1 绘制到上体部分,则破坏了跟踪对象的完整性,将提高误判的风险。此时,在保留 BB1 框定目标整体的基础上,在相对稳定的局部区域绘制 BB2,可以兼顾跟踪对象的“重点性”和“整体性”,从而提高跟踪精度。像这样的目标还有很多,比如行走的人、被局部遮挡的物体、舞动肢体的八爪鱼、分节的火箭等等。

以《小海龟大历险》中八爪鱼的镜头为例,说明双矩形框区域标定方法的功能。图2中(e)~(h)显示了八爪鱼在跟踪过程的变化情况,可见,八爪鱼的四肢比头部变化更为剧烈。图4对影片初始帧的图像构成进行解析,整体而言,图像包含背景区域和目标区域,目标区域又由稳定区域和不稳定区域构成。BB1 覆盖目标整体部分,BB2 只框定变化较为稳定的头部部分。

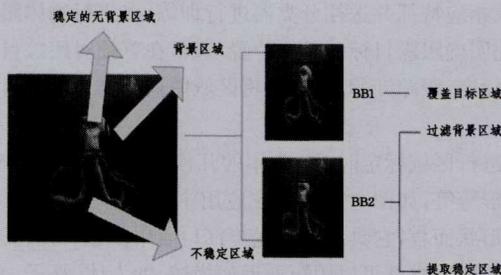


图4 双矩形框标定的八爪鱼

4 TLD 算法改进

TLD 算法在跟踪第一帧中的目标时,在标定的矩形框区

域内均匀采样 100 个点,用中值光流法进行跟踪^[6,12],得到其在下一帧的位置,并生成新的矩形框。算法中采取 NCC (Normal Cross-correlation, 归一化互相关系数) 和 Forward-Backward Error^[3] 来进行筛选,最后剩下不到 50% 的跟踪点,这对部分不稳定点和在背景区域取样的点产生一定的过滤作用^[3]。但是,原 TLD 算法对跟踪的稳定区域和背景区域、非稳定区域不作区分,以均等概率提取跟踪点,即相同面积的重点区域和背景区域产生相同数量的跟踪点,但实际上,这些区域对跟踪者来说意义或者重要程度是不同的,无差别对待显然不合理。

在双矩形框标定技术的基础上,我们采用分配权重的方法,使得相同面积的不同区域产生不同数量的跟踪点。设权重比为 $k:1$,即在非稳定区域取样 1 个点时,在相同面积的稳定区域取样 k 个点。 k 的最优值由实验决定, num 为待产生的跟踪点总数。 $numS_i$ 为图中对应 S_i 区域产生的跟踪点数量, S_i 为 S_i 区域的面积, x 为单位面积非稳定区域产生的点数,可列方程如下:

$$x(S_1 + S_2 + S_3 + S_4) + x \cdot k \cdot S_5 = num \quad (1)$$

$$x = \frac{num}{k \cdot S_5 + (S_1 + S_2 + S_3 + S_4)}$$

各个区域产生的点数计算如下:

$$numS_5 = \frac{num}{k \cdot S_5 + (S_1 + S_2 + S_3 + S_4)} \cdot k \cdot S_5 \quad (2)$$

$$numS_1 = \frac{num}{k \cdot S_5 + (S_1 + S_2 + S_3 + S_4)} \cdot S_1 \quad (3)$$

$$numS_2 = \frac{num}{k \cdot S_5 + (S_1 + S_2 + S_3 + S_4)} \cdot S_2 \quad (4)$$

$$numS_3 = \frac{num}{k \cdot S_5 + (S_1 + S_2 + S_3 + S_4)} \cdot S_3 \quad (5)$$

$$numS_4 = \frac{num}{k \cdot S_5 + (S_1 + S_2 + S_3 + S_4)} \cdot S_4 \quad (6)$$

如图5所示,(d)遵循原 TLD 算法,用单个矩形框进行区域标定,记作 BB1,(e)中用“X”作标记标出均匀取样的 100 个点,通过光流法跟踪预测,并利用 NCC 和 FB-error 进行过滤^[2],得到 26 个跟踪点,用“□”标识,如(f)所示;(a)中引入双矩形框,记作 BB1 和 BB2,BB1 覆盖整体,而 BB2 仅覆盖稳定区域。(b)中按新算法抽样取点,这里将权重比定为 1:1.5,(c)中用“□”标识跟踪后的 28 个点。对比(b)(e)和(c)(f)可见,(e)中均匀取点,由于背景区域和不稳定区域面积较大,应作为重点跟踪部分的稳定区域反而获得较少的跟踪点。在进行光流法跟踪和筛选之后,(f)中稳定区域剩余的点就更多了。而在(b)(c)中,稳定区域一直分布有较多的跟踪点。

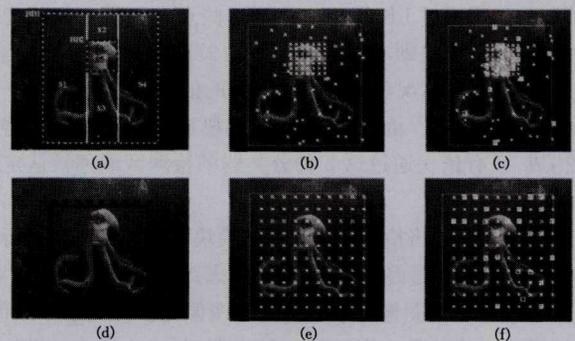


图5 TLD 算法改进前后跟踪点选取对照图

5 实验

为了与原 TLD 算法进行比较,实验在利用 Kalal 提供数据库^[13](下文称作标准数据库)的同时,又构造了爬行的熊、八爪鱼、教室行走的学生、讲课的老师、餐厅的服务员等数据库。文章第 3 节提到,在一些场景中,被跟踪目标会出现局部形态保持稳定,而其余部分持续变化的情况。此时,基于单个矩形框的标定的 TLD 算法无法兼顾跟踪对象的“重点性”和“整体性”,导致跟踪出现丢帧现象。新构造的数据库大多基于这样的场景,用来证明本文基于双矩形框标定算法的有效性和先进性。图 6 简单展示了 4 组数据库的目标标定和跟踪情况,(a)(b)为标准数据库,(c)(d)为新构造的数据库。每组显示两张图片,第一张为利用双矩形框标定的初始帧,另一张为跟踪过程中的某一帧。



图 6 初始帧标定和跟踪结果图示

实验选取 Precision(准确率,简称 P)、Recall(召回率,简称 R)、F-measure(综合评价指标,简称 F1)作为跟踪算法评估指标^[2],其中 $F1 = 2PR / (P + R)$,实验结果如表 1 所列。

表 1 改进前后 TLD 算法的实验结果

数据库	TLD		改进的 TLD 算法	
	P/R/F	P/R/F	P/R/F	P/R/F
Jumping	0.99/0.88/0.93	0.99/0.90/0.94	0.99/0.90/0.94	0.99/0.90/0.94
Motocross	0.92/0.32/0.47	0.90/0.68/0.77	0.90/0.68/0.77	0.90/0.68/0.77
Car Chase	0.90/0.42/0.57	0.85/0.40/0.55	0.85/0.40/0.55	0.85/0.40/0.55
Panda	0.51/0.16/0.25	0.71/0.63/0.67	0.71/0.63/0.67	0.71/0.63/0.67
八爪鱼	0.85/0.31/0.45	0.88/0.93/0.90	0.88/0.93/0.90	0.88/0.93/0.90
熊	0.92/0.35/0.50	0.93/0.70/0.79	0.93/0.70/0.79	0.93/0.70/0.79

从表 1 可见,除 Car Chase 数据库以外,改进后的算法对 Recall 均有提升,使得整体跟踪性能有所改观。在 Jumping 数据库中,人在跳动过程中,头部始终在摄像头内且形态变化不大,用矩形框标定头部可以捕捉跟踪对象的稳定特征。同时,虽然跳动时身体被遮挡面积在不断变化,但整体形态变化不大,角度基本没变,使得外框标定区域的特征也基本稳定。因此,算法改进前后效果提升不明显;在 MotorCross、Panda、八爪鱼、熊数据库中,与跟踪目标的其它区域相比,跟踪目标的稳定局部区域较为突出,且在较长的拍摄时间内保持较高的一致性。用矩形框将其标定,便捕捉了跟踪目标的关键部位减少了丢帧现象,提高了跟踪的召回率。

对于 Car Chase 数据库,算法性能反而降低。这是由于在车辆行驶中,跟踪目标各个部分都存在形变以及被遮挡的情况,即对于目标来说,没有固定的稳定局部区域。这使得双

矩形框标定中标定稳定区域的矩形框失去原有的效果,导致算法性能与基于单个矩形框标定的 TLD 算法相比,基本保持不变。

结束语 TLD 算法对于单目标的长期跟踪具有较好效果,但对于一些目标变化剧烈的场景来说,会出现丢帧现象。为提高算法性能,文章设计实现双矩形框标定方法,并利用其对基于单个矩形框标定的 TLD 进行改进。实验表明,改进后算法对包含稳定局部区域的目标有较好的跟踪效果,而对于不包含固定稳定局部区域的跟踪目标,跟踪效果没有明显改善。另外,由于矩形框需要手工标定,其尺寸和大小对跟踪效果的影响不容忽视。在下一步的研究工作中,把智能识别稳定区域作为研究的重点,实现矩形框的自适应绘制,以克服人工标定带来的不确定性。

参考文献

- [1] Wang Qing, Chen Feng, Xu Wen-li, et al. An experimental comparison of online object-tracking algorithms[C]//Proceedings of SPIE: Image and Signal Processing, 2011: 138-147
- [2] Zdenek K, Jiri M, Krystian M. Tracking-learning-detection[J]. IE-EE Trans on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1409-1422
- [3] Zdenek K, Jiri M, Krystian M. Forward-Backward Error: Automatic Detection of Tracking Failures[C]//International Conference on Pattern Recognition (ICPR). 2010: 23-26
- [4] Zdenek K, Jiri M, Krystian M. P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints[C]//Computer Vision and Pattern Recognition (CVPR). 2010: 49-56
- [5] 程立英,张丹,赵姝颖,等.一种基于 TLD 改进的视觉跟踪算法[J].科学技术与工程,2013,13(9):2382-2385
- [6] Lucas B D, Kanade T. An iterative image registration technique with an application to stereo vision[C]//Proceedings of Imaging Understanding Workshop. 1981: 121-130
- [7] 朱克忠.基于光流法对移动目标的视频检测与应用研究[D].合肥:合肥工业大学,2007
- [8] Tomas V. Long-term Visual Object Tracking with online Learning [D]. Czech Technical University, 2012
- [9] Wang K, Babenko B, Belongie S. End-to-End Scene Text Recognition[C]//2011 IEEE International Conference on Computer Vision (ICCV). 2011: 1457-1464
- [10] Rupali S C, Patil. Object Tracking Based on Tracking-Learning-Detection [OL]. http://gnebehay.github.io/OpenTLD/genbehay_thesis_msc.pdf
- [11] Li Tian-yu, Liu Jia, Gong chen, et al. Robust Object Tracking by Combining Boosting Learning and Particle Filter[J]. Journal of Computational Information Systems, 2012, 22(8): 9593-9601
- [12] Zdenek k, Krystian M, Jiri M. Face-TLD: Tracking-learning-detection applied to faces[C]//ICIP. 2010: 26-29
- [13] The standard TLDdata[OL]. <http://personal.ee.surrey.ac.uk/Personal/Z.Kalal/index.html>