

文章编号 1004-924X(2015)08-2339-10

用基于二值化规范梯度的跟踪学习检测 算法高效跟踪目标

程 帅¹, 曹永刚^{1,2}, 孙俊喜^{3*}, 刘广文¹, 韩广良²

- (1. 长春理工大学 电子信息工程学院, 长春, 130022;
2. 中国科学院 长春光学精密机械与物理研究所, 长春, 130000;
3. 东北师范大学 计算机科学与信息技术学院, 长春, 130117)

摘要:为提高复杂环境下 TLD(Tracking-Learning-Detection)算法的跟踪精度和速度,提出基于二值化规范梯度(BING)的高效 TLD 目标跟踪算法。在跟踪器中引入基于时空上下文的局部跟踪器失败预测方法和全局运动模型评估算法,提高了跟踪器准确度和鲁棒性;用 BING 算法取代滑动窗口搜索策略,结合级联分类器实现目标检测,减少了检测器的检测范围,提高了检测的处理速度;将训练样本权重整合到在线学习过程中,改进级联分类器的分类准确度,解决了目标漂移问题。对不同的图片序列实验结果表明:本算法的跟踪正确率达 85%,帧率达 19.79 frame/s。与原始 TLD 算法及其他主流跟踪算法相比较,该算法在复杂环境下具有更高的鲁棒性、跟踪精度及处理速度。

关键词:目标跟踪;跟踪-学习-检测;二值化规范梯度;加权

中图分类号:V556 文献标识码:A doi:10.3788/OPE.20152308.2339

Efficient target tracking by TLD based on binary normed gradients

CHENG Shuai¹, CAO Yong-gang^{1,2}, SUN Jun-xi^{3*}, LIU Guang-wen¹, HAN Guang-liang²

- (1. School of Electronic Information Engineering,
Changchun University of Science and Technology, Changchun 130022, China;
2. Changchun Institute of Optics, Fine Mechanics and Physics,
Chinese Academy of Sciences, Changchun 130000, China;
3. School of Computer Science and Information Technology,
Northeast Normal University, Changchun 130117, China)
* Corresponding author, E-mail: junxi_sun@126.com

Abstract: To improve the tracking precision and processing speed of the Tracking-Learning-Detection(TLD) algorithm under a complex environment, an efficient TLD target tracking algorithm based on Binary Normed Gradient(BING) algorithm was proposed. The local tracker failure predicting method based on spatial-temporal context and the global motion model estimation algorithm was introduced into the tracker to improve its precision and robustness. Then, the BING algorithm was used to replace a sliding window for searching the target to detect the candidate target by combining with a cascaded classifier, so that to reduce the search space and improve the processing speed of the detector. The sample weight was integrated into the online learning

收稿日期:2015-03-07;修订日期:2015-04-16.

基金项目:国家自然科学基金资助项目(No. 61172111);吉林省科技厅资助项目(No. 20090512, No. 20100312)

procedure to improve the accuracy of the classifier and to alleviate the drift to some extents. The experimental results on variant sequences demonstrate that the accurate rate and the frame rate of the improved TLD are 85% and 19.79 frame/s, respectively. Compared with original TLD and state-of-the-art tracking algorithm under the complex environment, the improved TLD has the superior performance on robustness, tracking precision and tracking speeds.

Key words: target tracking; Tracking-Learning-Detection(TLD); BInary Normed Gradient(BING); weighting

1 引言

目标跟踪是计算机视觉研究领域的重要组成部分,广泛应用在视频监控、行为识别及国防安全等领域^[1]。影响目标跟踪性能的因素很多,如目标外观改变、光照变化、遮挡、相似目标等^[2]。

基于目标外观模型可将跟踪算法分为生成模型^[3-6]和判别模型^[7-11]方法。生成模型法通过构建外观模型对目标进行描述,在后续每一帧图像中,将匹配度最高的目标作为跟踪结果;判别模型法则将跟踪视为二值分类问题,有效地利用目标和背景信息训练分类器,利用训练好的分类器把目标从背景中分离出来。在此类算法中,不精确的分类可能导致样本的错误标记,以错误标记的样本训练分类器必然降低分类器的性能。为解决上述问题,Boris 等人^[12]提出多示例跟踪算法,该算法不是对单个样本进行标注,而是对多样例包进行标注来达到弱化监督效应的目的。Zhang 等人^[13]提出的加权多示例目标跟踪算法根据样本距离目标位置的远近设置样本的权重,同时在学习过程中采用权值最大的样本训练分类器,从而提高分类的准确度。

Kalal 等人^[14]提出的跟踪-学习-检测(Tracking-Learning-Detection, TLD)跟踪算法将跟踪、检测、在线学习机制相结合,有效地解决了跟踪过程中目标形变、光照变化、部分遮挡等问题。但 TLD 算法存在以下问题:跟踪器预测过程中耗时且精度不高;检测器基于多尺度滑动窗口,此过程窗口数量大,且大部分窗口都为背景信息,因此增加了不必要的计算量;在线 P-N(Positive-Negative)学习过程中对无标记数据常出现错误标记,用错误标记的样本训练分类器,得到的是不精确的分类器。

文献^[15]、^[16]分别对 TLD 算法进行了改

进。文献^[15]利用单元格群跟踪器(Flock of Tracker, FoT)+ Σ 预测器^[17]提高跟踪器的跟踪速度与精度,同时采用 Kalman 区域预测减少窗口数量,但区域预测不适用于目标快速运动及目标消失后运动方向改变的情况。文献^[16]将局部搜索和全局搜索相结合,并采用积分直方图过滤器来减少搜索区域,利用随机 Harr-like 块特征过滤器提高分类器的分类精度,但当目标被遮挡或在视场中消失时,全局搜索策略依然很耗时。

在目标检测过程中,物体检测(objectness measure)算法^[18-20]可取代滑动窗口搜索策略,有效地检测物体区域,减少搜索空间,提高计算效率。与文献^[18-20]中的算法相比,程明明等人^[21]提出的 BING(BInary Normed Gradients)算法具有检测率高、通用性强、计算速度达 300 frame/s 等优点。

为解决 TLD 算法存在的问题,本文提出基于 BING 算法的物体区域检测 TLD 目标跟踪(BING TLD, BTLTD)算法,即对原始 TLD 算法的跟踪器、P-N 学习、检测器 3 个组件分别进行改进,从而实现更加稳定、可靠、快速的目标跟踪。

2 BTLTD 算法

BTLTD 算法由 4 部分构成:增强 FoT 跟踪器、检测器、加权 P-N 学习、整合器。其中,增强 FoT 跟踪器和检测器对每一帧图像进行并行处理;整合器用于整合跟踪结果和检测结果,得到下一帧中目标的最终位置,并将整合结果反馈给加权 P-N 学习,以验证数据的准确性;加权 P-N 学习用于实时更新增强 FoT 跟踪器和检测器的模型参数。图 1 为 BTLTD 算法框图,斜黑体部分为改进方法,每个组件的描述如下:

增强 FoT 跟踪器:预测相邻两帧目标的运动位置。在原有 FoT 跟踪器的基础上,利用级联预

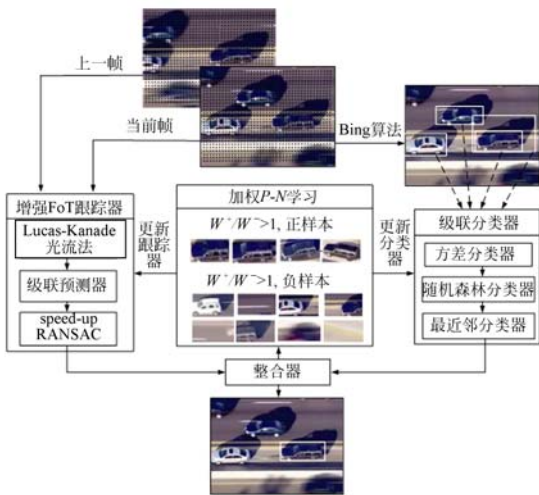


图 1 BTLD 算法框图

Fig. 1 Block diagram of BTLD algorithm

测器取代原有前向后向 (Forward Backward, FB) 误差预测器^[14]和归一化互相关 (Normalized Cross Correlation, NCC) 预测器^[14], 预测局部跟踪器是否跟踪成功。添加快速随机采样一致性 (RANdom SAmples Consensus, RANSAC) 算法^[22], 用以评估跟踪器全局运动模型。

检测器: 检测一帧中所有可能的目标, 并可以解决跟踪器不能跟踪快速运动目标及目标被遮挡或消失再出现在视场中重新定位问题。将 BING 算法取代滑动窗口策略, 搜索图像中物体区域。级联分类器^[14]对每个区域进行分类, 确定每个区域中是否含有目标。级联分类器由方差分类器、随机森林分类器、最近邻分类器构成。

加权 P-N 学习: 利用增长事件和裁剪事件评估检测误差^[23], 在线更新目标模型及跟踪器和检测器的相关参数。在 P-N 学习算法的基础上, 对训练样本赋予正负两类权值, 通过两类权值的比值决定样本是正样本还是负样本。

整合器: 将跟踪器的预测信息和检测器的检测信息进行整合, 用于判断当前帧是否存在目标、目标位置信息及当前帧的跟踪轨迹是否有效等信息。

2.1 增强 FoT 跟踪器

在原有 FoT 跟踪器的基础上, 利用局部跟踪器时空上下文信息增添邻域一致预测器、马尔科夫预测器, 并与 NCC 预测器构成级联预测器。快速 RANSAC 算法利用跟踪成功的局部跟踪器评估全局运动模型, 预测目标的位置。

2.1.1 级联预测器

在级联预测器中, NCC 预测器通过求取跟踪所得图像区域与之前被跟踪区域之间的相似程度来判断局部跟踪器是否有效。马尔科夫预测器利用时间上下文信息构建马尔科夫链模型, 根据当前时刻的信息预测下一时刻局部跟踪器正确跟踪的概率, 判断局部跟踪器是否跟踪成功。邻域一致预测器假定短时间内相邻局部跟踪器之间具有运动一致性, 利用空间上下文信息通过相邻跟踪器预测局部跟踪器是否跟踪成功, 具体实现可参考文献^[17]。级联预测器预测框图如图 2 所示。3 种预测器间为级联关系, 每种预测器对局部跟踪器是否跟踪成功进行判断, 局部跟踪器只有符合 3 种预测器的条件才认定为跟踪成功, 否则跟踪失败。

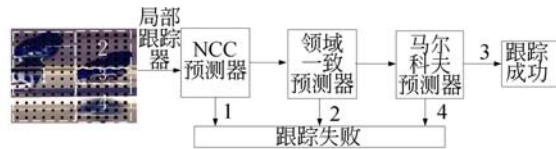


图 2 级联预测器框图

Fig. 2 Block diagram of cascaded predictor

2.1.2 快速 RANSAC 算法

Botterill 等人^[22]提出的快速 RANSAC 算法能够从包含大量外点的数据集中快速、准确地估计出最优参数模型, 使用贝叶斯算法更新内点集合的概率, 减少迭代次数, 降低模型失败的概率。根据级联预测器得到的局部跟踪器集合, 采用快速 RANSAC 算法评估目标最优全局运动模型, 可以在增加很少计算成本的前提下, 提高跟踪器对目标位置的预测精度, 解决目标漂移问题。

2.2 BING 算法

将包含物体与非物体的窗口归一化到 8×8 大小, 计算梯度特征, 得到的 64 维梯度特征称为规范梯度 (Normed Gradient, NG) 特征, 每一点梯度均归一化到 $[0, 255]$ 之间, 物体与非物体之间的 64 维梯度特征有着明显的差别。

利用物体与非物体之间 NG 特征的差别, BING 算法使用级联支持向量机 (Support Vector Machine, SVM) 分类器检测图像中的物体^[24]。首先将原图像缩放到预定义大小 (36 种不同尺寸), 计算其梯度特征, 利用 8×8 滑动窗口扫描缩放后的图像, 利用第一级 SVM 线性模型 w 计算

所有窗口的滤波分数,分数越高包含物体的概率越大,计算公式为:

$$s_l = \langle \mathbf{w}, g_l \rangle, \quad (1)$$

$$l = (i, x, y), \quad (2)$$

其中: s_l 为滤波分数; g_l 为 NG 特征; l 为窗口位置; i 为预定义尺度; (x, y) 为窗口内坐标。

使用非极大抑制算法在每一预定义尺寸下选择部分窗口作为包含物体的候选窗口^[25]。由于不同大小的窗口包含类物体的概率也是不同的,使用第二级 SVM 线性模型计算所有候选窗口的物体分数,每一尺寸下物体分数的计算公式如下:

$$o_i = v_i \cdot s_l + t_i, \quad (3)$$

其中: i 为预定义尺寸; v_i 为线性模型; t_i 为偏差项。

最后根据物体分数对窗口进行排序,选取分数高的窗口作为物体区域,并将这些窗口输入到级联分类器进行分类。

为加快计算速度,二值化处理线性模型 \mathbf{w} 。用一组基向量的线性组合表示 \mathbf{w} ,其计算公式如下:

$$\mathbf{w} = \sum_{j=1}^{N_w} \beta_j \boldsymbol{\alpha}_j, \quad (4)$$

其中: $\mathbf{w} \in \mathcal{R}^{64}$, $\boldsymbol{\alpha}_j \in \{-1, 1\}^{64}$ 为基向量, $\beta_j \in \mathcal{R}$ 为相关系数。

每个基向量可用二值向量和它的补向量表示,计算公式如下:

$$\boldsymbol{\alpha}_j = \boldsymbol{\alpha}_j^+ - \overline{\boldsymbol{\alpha}_j^+}, \quad (5)$$

其中: $\boldsymbol{\alpha}_j^+ \in \{0, 1\}^{64}$ 。

为加快特征提取速度,对 NG 特征二值化处理。 8×8 大小的 NG 特征中每个点梯度值可由 8 位二进制数表示,NG 特征为 $8 \times 8 \times 8$ 矩阵,每个矩阵元素为二进制数。每次读取 64 维(NG 特征的 8×8) 二进制数,共 8 次,完成 NG 特征的提取。读取 64 维二进制过程中每 8 个为一组,移位 8 次,此过程仅为移位运算,大大提高了计算效率。为节省存储空间,每一点的梯度特征仅有高 N_g 位近似表示。二值化处理后的 NG 特征称为 BING 特征, g_l 可近似表示:

$$g_l = \sum_{k=1}^{N_g} 2^{8-k} b, \quad (6)$$

其中, b 为 64 维二进制数。

2.3 加权 P-N 学习

在线 P-N 学习算法^[9]在对样本标记过程中常出现错误标记问题,若用错误标记的样本训练

分类器会很大程度上降低分类器性能,导致目标漂移甚至跟踪失败。为解决上述问题,在分类过程中,对每个训练样本赋予两类权值:正样本权值 W^+ 、负样本权值 W^- 。正、负样本权值分别由两部分构成:

$$\begin{aligned} W^+ &= W_b^+ + \eta W_c^+ \\ W^- &= W_b^- + \eta W_c^- \end{aligned}, \quad (7)$$

其中: W_b 为 bootstrapping 分类过程赋予的权值, W_c 为随机森林分类过程赋予的权值。+ 表示正样本, - 表示负样本, η 为平衡系数。

在 bootstrapping 分类过程中, W_b^+ 和 W_b^- 的计算公式如下:

$$\begin{aligned} W_b^+ &= f^+ / (f^+ + f^-) \\ W_b^- &= f^- / (f^+ + f^-) \end{aligned}, \quad (8)$$

其中: f^+ 、 f^- 分别为样本被分类为正样本、负样本的次数。

随机森林分类器对每个包含物体的窗口进行分类,每个窗口的特征向量为 \mathbf{x} ,每个窗口被分类为正样本的后验概率为 $P(y=1 | \mathbf{x})$,则分类权值的计算公式为:

$$\begin{aligned} W_c^+ &= P(y=1 | \mathbf{x}) \\ W_c^- &= 1 - P(y=1 | \mathbf{x}) \end{aligned}. \quad (9)$$

样本最终被分类为正样本、负样本的计算公式如下:

$$\begin{aligned} W^+ / W^- &\geq 1, \text{ 正样本} \\ W^+ / W^- &< 1, \text{ 负样本} \end{aligned}. \quad (10)$$

3 实验结果与分析

实验共分为 2 部分:第 1 部分为 BTLD 算法与原始 TLD 算法中跟踪器、检测器与学习算法的对比实验;第 2 部分为 BTLD 算法与目前主流跟踪算法的对比实验。试验中选取 5 组图片序列^[14]: pedestrian2、pedestrian3、car、carchase、panda。表 1 为 5 组图片序列各自的特点。为了

表 1 图片序列特点

Tab. 1 Features of image sequences

图片序列	总帧数	特点
pedestrian2	338	遮挡、相似目标
pedestrian3	184	遮挡、旋转、相似目标
car	945	遮挡、相似目标
carchase	9 928	遮挡,相似目标,光照、尺度变化
panda	3 000	遮挡,姿态、光照、尺度变化

使结果更加一般化,每个算法在每组序列上均进行 20 次测试,将 20 次结果的平均值作为最终结果。

实验 1 首先对跟踪器进行性能比较,设 $T_{\text{FB+NCC}}$ 为原始 TLD 中的跟踪器, T_{cascade} 为只有级联预测器的跟踪器, T_{cascade^+} 为具有级联预测器和快速 RANSAC 算法的跟踪器,通过比较成功跟踪帧数对跟踪器进行评估,跟踪结果与真实位置的重叠度大于 0.5,则认定为跟踪成功,表 2 为 3 种跟踪器成功跟踪帧数。

表 2 目标成功跟踪帧数

Tab. 2 Frame number of successfully tracking targets

图片序列	目标可见帧数	$T_{\text{FB+NCC}}$	T_{cascade}	T_{cascade^+}
pedestrian2	266	100	235	263
pedestrian3	156	52	52	60
car	860	510	580	640
carchase	8 660	182	231	301
panda	2 730	70	95	124

表 2 第 3 列为 $T_{\text{FB+NCC}}$ 的正确跟踪帧数,由表 2 可见, $T_{\text{FB+NCC}}$ 正确跟踪帧数少于其他两种跟踪器的跟踪帧数,因为 $T_{\text{FB+NCC}}$ 中利用 FB 预测器和 NCC 预测器判断局部跟踪器是否跟踪正确,FB 预测器仅根据跟踪轨迹先后一致性假设对局部跟踪器进行预测,即跟踪轨迹与验证轨迹之间的欧式距离小于 10 pixel,则认定为跟踪成功,反之跟踪失败。同时正确的跟踪不依赖于时间流的方向,没有考虑时间上下文信息,因此预测局部跟踪器是否正确跟踪的准确度会下降,跟踪失败的局部跟踪器增多将导致跟踪器总体性能下降,成功跟踪帧数将减少。

表 2 第 4 列为 T_{cascade} 跟踪结果,与 $T_{\text{FB+NCC}}$ 的跟踪结果相比,正确跟踪帧数明显增加,因为 T_{cascade} 中新添加了邻域一致预测器和马尔科夫预测器,并与 NCC 预测器构成级联预测器。其中:邻域一致预测器利用局部跟踪器的空间上下文信息判断局部跟踪器是否成功,而马尔科夫预测器利用时间上下文信息构成马尔科夫模型判断局部跟踪器是否成功。同时预测器间采用逻辑与的级联关系,只有符合 3 种预测器判定条件的局部跟

踪器才认定为跟踪成功,因此,提高了失败预测的成功率,减少了跟踪失败的局部跟踪器对跟踪结果的影响。

表 2 的第 5 列为 T_{cascade^+} 跟踪结果,其正确跟踪帧数高于其他两种跟踪器的结果。 T_{cascade^+} 在 T_{cascade} 的基础上新融合了快速 RANSAC 算法,级联预测器完成对局部跟踪器预测后,舍弃没有正确跟踪的局部跟踪器,快速 RANSAC 算法利用正确跟踪的局部跟踪器集合评估最优全局运动模型,即使有失败局部跟踪器的干扰,快速 RANSAC 算法依然可以最大程度上得到最优的运动模型,有效解决目标漂移问题,实现更鲁棒的目标跟踪。

表 3 中给出了 3 种短时跟踪器跟踪每一帧的平均处理速度。由表 3 可知, $T_{\text{FB+NCC}}$ 的处理速度最慢, T_{cascade} 的处理速度最快,这是因为 $T_{\text{FB+NCC}}$ 大部计算时间都消耗在 FB 预测器的预测过程,FB 预测器利用 LK(Lucas-Kanade) 光流法要得到跟踪轨迹和验证轨迹,因此,需要计算两次 LK 光流法,而 T_{cascade} 新添加的邻域一致预测器和马尔科夫预测器只需计算一次 LK 光流算法就可以完成对局部跟踪器的预测,不但提高了跟踪的鲁棒性,而且减少了计算成本。 T_{cascade^+} 的计算时间多于 T_{cascade} 的部分用在快速 RANSAC 算法对全局运动模型的评估上,但快速 RANSAC 算法将贝叶斯算法与 RANSAC 算法相结合,减少了评估最优模型的迭代次数,从而提高了计算速度,其计算时间明显少于一次 LK 光流法的计算时间。

表 3 跟踪每一帧所用平均时间

Tab. 3 Average consuming time of tracking a frame

跟踪器	$T_{\text{FB+NCC}}$	T_{cascade}	T_{cascade^+}
时间/ms	16.64	11.52	13.23

其次,对检测器进行性能比较, D_{scan} 为基于滑动窗口的检测器, D_{BING} 为基于 BING 算法的检测器,两者都是使用级联分类器。通过成功检测帧数和计算时间评估检测器的性能,检测位置与真实位置的重叠度大于 80%,则认为检测成功,表 4 为成功检测帧数。

表 4 目标成功检测帧数

Tab. 4 Number of successfully detecting targets

图片序列	目标可见帧数	D_{scan}	D_{BING}
pedestrian2	266	245	242
pedestrian3	156	156	156
car	860	834	830
carchase	8 660	6 011	6 024
panda	2 730	1 719	1 719

由表 4 可知,除 carchase 序列外, D_{scan} 的成功检测帧数大于 D_{BING} 的。这是因为 TLD 中的检测器采用的是全局多尺度滑动窗口搜索策略,即穷举式检测图像大部分区域,而基于 BING 的检测器只得到包含物体区域的图像窗口,检测区域明显减少,导致成功检测帧数减少,但从结果来看相差帧数很小。

在 carchase 序列中, D_{BING} 的检测帧数大于 D_{scan} 检测帧数,由于两个检测器中仅采用像素级特征对目标进行描述,把与模板匹配度最高的窗口作为最终的跟踪结果,当有相似背景干扰的情况下,常将背景作为检测结果,因此出现错误检测问题。与基于 BING 算法相比,滑动窗口扫描过程中会产生更多的包含背景信息的子窗口,因此在相似背景条件下增加了错误检测的机率。表 5 为两种检测器检测 carchase 序列每一帧所用的平均时间。

表 5 检测每一帧所用平均时间

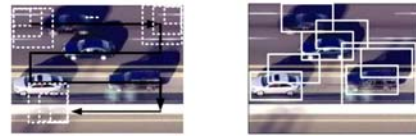
Tab. 5 Average consuming time of detecting a frame

检测器	D_{scan}	D_{BING}
时间/ms	18.61	11.42

由表 5 可知, D_{scan} 的计算时间长于 D_{BING} 的计算时间。这是由于对目标的搜索策略不同导致的。图 3 为滑动窗口与 BING 算法的采样过程。如图 3(a)所示,基于滑动窗口的检测器,滑动窗口的放大因子为 $s=1.2^b$, $b \in \{-10, -9, \dots, 9, 10\}$,子窗口在 x 轴和 y 轴的移动步长分别为输入图片高和宽的 10%,一幅大小为 $240 \text{ pixel} \times 320 \text{ pixel}$ 的图像,可产生 50 万多个子窗口。检测器需要对每个子窗口内的图像数据进行分类以

确定是否含有目标,而且大多数子窗口不包含目标信息,从而增加了分类器不必要的计算量。

BING 算法首先采样图片中所有物体的区域,大约得到 2 000 个左右物体区域窗口,如图 3(b)所示,图中仅显示了 8 个物体中分数最高的窗口,跟踪目标包含在其中(目标为右下角黑色小车),再通过级联分类器决定哪个窗口区域为最终的目标。此方法在满足精度需求的前提下,很大程度上减少了检测区域,提高了检测速度。



(a) 滑动窗口 (b) BING
(a) Sliding window (b) BING

图 3 窗口采样过程

Fig. 3 Window sampling process

最后,对加权 P-N 学习算法和原始 P-N 学习算法进行性能比较,跟踪器与检测器为 TLD 算法中的原始部分,图 4 为对比结果。



(a) car



(b) carchase



(c) panda

图 4 加权 P-N 学习跟踪结果

Fig. 4 Tracking results with weighted P-N learning

在图 4(a)、(b)、(c)中,最左侧一列为目标的真实位置,中间一列为 P-N 学习跟踪结果。从图中可以看出,P-N 学习跟踪结果出现了目标漂移问题,因为当分类过程中出现错误标记时,用错误

标记的样本训练分类器会降低分类器分类性能,导致目标漂移。最右侧一列为加权 P-N 学习跟踪结果,与 P-N 学习跟踪结果相比,加权 P-N 学习将样本权重整合到在线学习过程中,根据正、负样本权值决定样本最后的标记,在一定程度上解决了分类器错误标记问题,提高了分类器分类精确度,解决了目标漂移问题。

实验 2 与 BSBT^[26]、MIL^[8]、coGD^[27]、TLD 跟踪算法进行对比,这些算法均使用原作者提供的默认参数。通过正确率和帧率来评估算法的整体性能。正确率为正确跟踪帧数与跟踪总帧数的比值,逐帧计算跟踪算法得到的跟踪结果与真实位置的重叠度大于 0.5,则认为跟踪正确。表 6 给出各跟踪算法在不同图片序列的正确率和帧率的测试结果。

由表 6 可知,最后 1 行(mean 行)为测试结果的平均值,BTLD 算法的平均正确率为 0.85,帧率为 19.79 frame/s。BTLD 算法在每个图片序列的正确率和帧率均高于其他算法。BTLD 中级联预测器对局部跟踪器的失败性预测更准确,快速 RANSAC 算法通过评估最优运动模型,可以实现准确的目标位置预测,一次 LK 光流法的计算时间与快速 RANSAC 算法的评估时间要短于两次 LK 光流法的计算时间,因此提高了跟踪速度。改进 P-N 算法可解决分类过程中错误标记问题,解决目标漂移问题,提高了目标跟踪的准确率。相比于滑动窗口搜索策略,BING 算法在保证检测精度基本不变的前提下,减少了检测范围,提高了检测速度。

表 6 测试结果

Tab. 6 Sequence test results

图片序列	目标可见数	正确率					帧率/frame · s ⁻¹				
		BSBT	MIL	coGD	TLD	BTLD	BSBT	MIL	coGD	TLD	BTLD
pedestrian2	266	0.37	0.23	0.71	0.89	0.91	7.83	9.21	5.6	19.46	22.20
pedestrian3	156	0.81	0.49	0.84	0.99	1.00	5.64	6.52	4.1	15.33	19.29
car	860	0.56	0.11	0.91	0.92	0.95	8.56	10.23	4.6	14.66	19.27
carchase	8 660	0.38	0.49	0.73	0.50	0.86	8.27	7.10	8.10	13.43	17.62
panda	2 730	0.44	0.14	0.16	0.32	0.51	6.31	8.32	7.81	14.22	20.56
mean	1 985	0.51	0.29	0.67	0.72	0.85	7.32	8.26	6.04	15.42	19.79



(a) pedestrian2



(b) pedestrian3



(c) car



(d) carchase



(e) panda

图 5 不同算法跟踪结果对比

Fig. 5 Comparison of different tracking algorithms

图 5 为各算法在 5 组图片序列的跟踪结果。

pedestrian2 序列中跟踪目标是运动的行人,在跟踪过程中伴有摄像头快速运动、遮挡和相似背景的干扰。在第 41 帧,摄像头的快速运动导致 MIL 算法跟踪失败,而其他算法均可正确跟踪。在 82、93 帧中存在摄像头快速运动和目标被部分遮挡的情况,此时,只有 BTLD、TLD 算法能够正确跟踪,其他算法都已跟踪失败。在 164、180 帧,目标从视场中重新出现,此时,只有 BTLD 算法能够重新定位到目标,其他算法均跟踪失败,在相似背景的干扰下,coGD 算法出现了错误跟踪。可见,在快速运动、遮挡及相似背景环境下,BTLD 算法能够实现稳定的目标跟踪。

pedestrian3 序列中行人在行走,跟踪过程中伴有摄像头快速运动会导致模糊、目标在视场消失、遮挡和相似背景的干扰。在第 81、83、85 帧,摄像头快速运动导致目标在视场中消失,之后目标又重新出现,此时,BSBT、MIL 算法跟踪失败。在 116、120 帧中出现由快速运动引起的模糊拖影等现象,只有 BTLD 和 TLD 算法能够正确跟踪,其他算法均出现了目标漂移现象。可见,在快速运动、模糊拖影及相似背景下,BTLD 算法能够实现正确跟踪。

car 序列中小车在行驶过程中不断受到路边树木的遮挡,同时有相似目标的干扰,如 243 帧中 MIL 算法跟踪失败。由于目标受到树木的遮挡又再次出现在视场中,此时,只有 BTLD 算法能够稳定跟踪,其他算法均出现不同程度的漂移,甚

至跟踪失败,如 569、690、791 帧。由此可得 BTLD 算法可有效处理遮挡和相似背景问题,实现稳定跟踪。

对于 carchase 序列中快速行驶的小车,在跟踪过程中不断出现遮挡、尺度变化、光照变化和相似目标的干扰。在 175 帧,由于相似目标干扰,BSBT 算法出现错误跟踪,其他几种算法都有不同程度的目标漂移问题,只有 BTLD 能正确跟踪。目标消失后重新出现在视场中,同时伴有尺度变化时,只有 TLD 和 BTLD 算法能够重新定位到目标,但 TLD 算法有目标漂移现象,其他算法均跟踪失败,如 601、691 帧。在 1279 帧,由于相似目标的干扰,导致 BSBT 算法跟踪错误,只有 BTLD 算法能够正确跟踪。可见,BTLD 算法在遮挡、尺度变化、光照变化和相似目标条件下均能实现鲁棒的跟踪。

panda 序列中,大熊猫在走动过程中,不断出现姿势变化、旋转、遮挡、消失在视场中。由于目标姿态不断变化、旋转、倾斜和摇摆,导致了除 BTLD 算法以外的算法均出现目标偏移甚至跟踪失败,如 607、1448、1718、1819、帧。在 1330 帧,目标消失后重新出现在视场中,只有 BTLD 算法能够正确定位到目标,其他算法均跟踪失败。因此 BTLD 算法对目标姿态变化、旋转、遮挡具有很好的鲁棒性。

4 结 论

为提高 TLD 算法的跟踪速度与精度,本文

提出了基于 BING 算法物体区域采样的 TLD 目标跟踪——物体区域采样算法(BING 算法),用其取代传统的全局滑动窗口搜索策略,并应用到目标跟踪中,具有一定的实用价值。实验结果表明本算法提高了跟踪器的跟踪精度和速度,在保证检测精度的基础上,提高了检测器的处理速度,将样本权重整合到在线学习过程中,解决了目标

漂移问题,正确率为 0.85,帧率为 19.79 frame/s。本文提出的算法是基于物体区域采样的,只能跟踪物体目标,而且在相似背景条件下易出现错误跟踪,目前仅应用在单个目标跟踪中。下一步研究工作拟利用深度学习实现目标的有效表达,解决相似背景下错误跟踪问题,同时将该算法应用到多目标跟踪中。

参考文献:

- [1] 郭敬明,何昕,魏仲慧. 基于在线支持向量机的 Mean Shift 彩色图像跟踪[J]. 液晶与显示, 2014, 29(1): 120-128.
- GUO J M, HE X, WEI ZH H. New mean shift tracking for color image based on online support vector machine [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(1): 120-128. (in Chinese)
- [2] WU Y, LIM J, YANG M H. Online object tracking: A benchmark [C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, OR, Portland: IEEE, 2013; 2411-2418.
- [3] ROSS D A, LIM J, LIN R S, et al.. Incremental learning for robust visual tracking [J]. *International Journal of Computer Vision*, 2008, 77(1-3): 125-141.
- [4] 李静宇,王延杰. 基于子空间的目标跟踪算法研究[J]. 液晶与显示, 2014, 29(4): 617-622.
- LI J Y, WANG Y J. Subspace based target tracking algorithm [J]. *Chinese Journal of Liquid Crystals and Displays*, 2014, 29(4): 617-622. (in Chinese)
- [5] LI H, SHEN C, SHI Q. Real-time visual tracking using compressive sensing [C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI: IEEE, 2011:1305-1312.
- [6] SEVILLA-LARA L, LEARNED-MILLER E. Distribution fields for tracking [C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI: IEEE, 2012:1910-1917.
- [7] GRABNER H, GRABNER M, BISCHOF H. Real-time tracking via on-line boosting [C]. *Proceedings of British Machine Vision Conference*, Edinburgh, Scotland: BMVA, 2006, 47-56.
- [8] GRABNER H, LEISTNER C, BISCHOF H. Semi-supervised on-line boosting for robust tracking [C]. *Proceedings of European Conference on Computer Vision*, Berlin, Germany: Springer, 2008; 234-247.
- [9] KALAL Z, MATAS J, MIKOLAJCZYK K. P-N learning: Bootstrapping binary classifiers by structural constraints [C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, New York, USA: IEEE, 2010; 49-56.
- [10] STALDER S, GRABNER H, VAN G L. Beyond semi-supervised tracking: tracking should be as simple as detection, but not simpler than recognition [C]. *Proceedings of International Conference on Computer Vision Workshop*, Kyoto, Japan, 2009, 1409-1416.
- [11] 陈东成,朱明,高文,等. 在线加权多示例学习实时目标跟踪[J]. 光学精密工程, 2014, 22(6): 1661-1667.
- CHEN D CH, ZHU M, GAO W, et al.. Real-time object tracking via online weighted multiple instance learning [J]. *Opt. Precision Eng.*, 2014, 22(6): 1661-1667. (in Chinese)
- [12] BABENKO B, YANG M H, BELONGIE S. Robust object tracking with online multiple instance learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1619-1632.
- [13] ZHANGA K, SONG H. Real-time visual tracking via online weighted multiple instance learning [J]. *Pattern Recognition*, 2013, 46(1): 397-411.
- [14] KALAL Z, MIKOLAJCZYK K, MATAS J. Tracking-learning-detection [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(7): 1409-1422.
- [15] 周鑫,钱秋朦,叶永强,等. 改进后的 TLD 视频目标跟踪方法[J]. 中国图象图形学报, 2013, 18(9): 1115-1123.

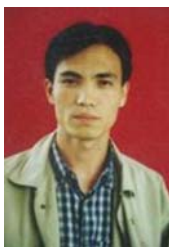
- ZHOU X, QIAN Q M, YE Y Q, *et al.*. Improved TLD visual target tracking algorithm [J]. *Journal of Image and Graphics*, 2013, 18(9):1115-1123. (in Chinese)
- [16] 江伟坚, 郭躬德. 复杂环境下高效物体跟踪级联分类器[J]. *中国图象图形学报*, 2014, 19(2): 253-265.
- JIANG W J, GUO G D. Efficient cascade classifier for object tracking in complex conditions [J]. *Journal of Image and Graphics*, 2014, 19(2): 253-265. (in Chinese)
- [17] VOJIR T, MATAS J. Robustifying the flock of trackers [C]. *Proceedings of Computer Vision Winter Workshop, Graz, Austria*, 2011: 91-97.
- [18] ENDRES I, HOIEM D. Category independent object proposals [C]. *Proceedings of European Conference on Computer Vision, Berlin, Germany*, 2010: 575-588.
- [19] ALEXE B, DESELAERS T, FERRARI V. Measuring the objectness of image windows [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012; 34(11): 2189-2202.
- [20] UJILING S J, VAN D S, GEVERS T, *et al.*. Selective search for object recognition [J]. *International Journal of Computer Vision*, 2013, 104(2): 154-171.
- [21] CHENG M M, ZHANG Z M, LIN W Y. BING: binarized normed gradients for objectness estimation at 300fps [C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Philip, Torr; IEEE*, 2014.
- [22] BOTTERILL T, MILLS S, GREEN R D. New conditional sampling strategies for speeded-up RANSAC [J]. *British Machine Vision Conference*, 2009: 1-11.
- [23] KALAL Z, MATAS J, MIKOLAJCZYK K. On-line learning of robust object detectors during unstable tracking [C]. *Proceedings of IEEE International Conference on Computer Vision Workshop, Kyoto, Japan; IEEE*, 2009: 1417-1424.
- [24] ZHANG Z, WARRELL J, TORR P H. Proposal generation for object detection using cascaded ranking SVMs [C]. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI; IEEE*, 2011: 1497-1504.
- [25] NEUBECK A, VAN G L. Efficient non-maximum suppression [C]. *Proceedings of International Conference on Pattern Recognition, Hong Kong, China*, 2006: 850-855.
- [26] YU Q, DINH T B, MEDIONI G. Online tracking and reacquisition using co-trained generative and discriminative trackers [C]. *Proceedings of European Conference on Computer Vision, Marseille, France*, 2008: 678-691.
- [27] STALDER S, GRABNER H, GOOL L V. Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition [C]. *Proceedings of International Conference on Computer Vision Workshops, Kyoto, Japan; IEEE*, 2009: 1409-1416.

作者简介:



程 帅(1987—),男,内蒙古呼伦贝尔人,博士研究生,2009年于长春工业大学获得学士学位,2012年于长春理工大学获得硕士学位,主要从事图像处理、目标跟踪、深度学习等方面的研究。
E-mail: chengshuai_pd@126.com

导师简介:



孙俊喜(1971—),男,河北唐山人,博士,教授,2004年于上海交通大学生物医学工程专业获得工学博士学位,2007年于中国科学院长春光学精密机械与物理研究所光学工程专业完成博士后研究,主要从事快速运动目标的检测与跟踪、嵌入式车牌识别系统、医学图像处理与分析。E-mail: junxi_sun@126.com