

基于在线学习的目标跟踪方法研究*

齐志泉, 宋野, 王来生[†]

(中国农业大学理学院数学系, 北京 100083)

摘要: 针对视频目标跟踪问题, 提出了一种基于 co-training 框架下的在线学习跟踪方法。该方法首先根据两种不同的局部特征, 利用在线 Boosting 算法分别建立模型, 然后采用 co-training 框架来协同训练, 有效避免了模型误差累积和跟踪丢帧等问题。实验证明了该方法的有效性。

关键词: 局部特征; 在线 Boosting; 协同训练; 目标跟踪

中图法分类号: TP18; O221 文献标志码: A 文章编号: 1001-3695(2010)02-0770-02

doi:10.3969/j.issn.1001-3695.2010.02.103

Object tracking research based on on-line learning

QI Zhi-quan, SONG Ye, WANG Lai-sheng[†]

(Dept. of Mathematics, College of Science, China Agricultural University, Beijing 100083, China)

Abstract: To video object tracking problem, this paper proposed an on-line learning tracking method based on co-training framework. First of all, the method adopted two different local features to build on-line Boosting model, and then, would train samples making use of co-training learning framework, which avoided the cumulative error of the model and dropping frames problem effectively. Furthermore, some experiments have been maded and the results implied that the new method is very efficient.

Key words: local features; on-line Boosting; co-training; object tracking

0 引言

视频目标跟踪是机器视觉领域的一个重要分支, 被广泛用于军事、多媒体检索、安全监控等领域。近年来, 将目标跟踪问题视为一个分类问题 (classification) 的方法是一个新的研究方向。主要思想 (图 (1)) 如下: a) 确定跟踪目标; b) 在跟踪目标区域随机扰动, 生成正类样本, 在背景区域生成负类样本; c) 进行目标分类, 得到目标跟踪分类器; d) 利用当前的目标分类器预测下一帧的图像, 得到目标位置, 重复 b)。

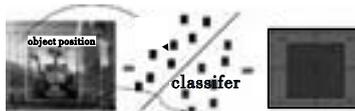


图1 目标跟踪的分类问题描述

根据上述思想进行目标跟踪的文献很多, 如 2005 年, Avitan^[1] 使用支持向量机对目标进行了跟踪; 同年, Lepetit 等人^[2] 以关键点 (key points) 作为特征, 利用随机树对目标进行跟踪; 随后 Grabner 等人^[3-5] 利用在线学习方法 (on-line Boosting) 对目标进行了有效跟踪。从实验结果来看, Grabner 等人的跟踪方法要优于前两者, 主要原因是采用了在线学习分类器 (on-line Boosting) 和局部特征 (HAAR、LBP) 提取等方法, 不但加快了模型更新速度, 而且提升了目标跟踪的准确率。

上述方法都属于利用分类器本身得到训练本来训练自身模型的过程, 属于自学习 (self-training) 方法。这种训练方法很容易累积误差, 使得分类器精度下降, 最后导致跟踪失败。Co-training 学习框架是解决上述问题的主要途径。该框架采

用 multi-view 的思想, 从两种互相独立的特征空间中提取特征, 然后利用两个分类器协同进行训练, 得到最终的分器, 如图 2 所示。

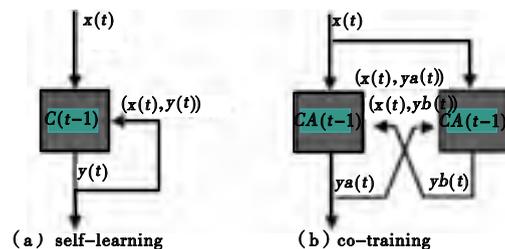


图2 自学习与协同训练

Co-training 学习框架最早由 Blum 等人^[6] 提出, 随后该方面的应用文章相继出现, 如 Collins 等人^[7] 将该框架用于文本分类, Javed 等人^[8] 利用该框架进行视频目标检测等。2009 年, 清华大学的王路等人^[9] 将 co-training 用于目标跟踪。在该框架中, 他们采用全局特征 (区域颜色直方图和区域梯度直方图) 和 off-line AdaBoost 算法。这样做的缺点如下:

- a) 计算全局特征比较耗时, 并且未对全局特征进行特征提取等降维处理;
- b) 颜色特征的选择限制了该算法的应用范围;
- c) Off-line AdaBoost 算法不能对样本执行依次输入, 不能及时调整特征跟踪模型。

本文提出一种基于在线学习的视频目标跟踪方法。该方法首先根据两种不同的局部特征 (Haar 特征和局部梯度直方图特征), 利用 on-line Boosting 算法进行特征选择并建立模型; 然后采用 co-training 框架来协同训练, 有效避免了模型误差累积和跟踪

收稿日期: 2009-06-15; 修回日期: 2009-07-27 基金项目: 国家自然科学基金资助项目 (60573158, 10771213)

作者简介: 齐志泉 (1979-), 男, 河北廊坊人, 博士研究生, 主要研究方向为模式识别、视频目标跟踪; 宋野 (1979-), 男, 辽宁沈阳人, 博士研究生, 主要研究方向为人脸识别、图像检测; 王来生 (1952-), 男 (通信作者), 教授, 博士, 主要研究方向为机器视觉、模式识别 (wanglaish@sina.com)。

丢帧等问题。尽管文献[8]也是 co-training 学习框架与 on-line Boosting 有效结合来进行目标检测,但采用的是 Oza^[10] 的 on-line Boosting 算法,只能取图像的全局特征,不能对图像的局部特征进行特征选择。本文采用的是 Helmut Grabner 的 on-line Boostings 算法,有效地解决了上述问题。

1 基于 multi-view 思想的特征选择

Co-training 学习框架采用的是 multi-view 思想,利用输入样本中互相独立的两种特征来训练两个互相独立的分类器。随后,这两个分类器将各自置信度较高的标记样本传输给对方来更新训练分类器,并不断重复这个过程。

对于选择特征而言,全局特征比较耗时,不符合实时跟踪的要求,因此本文采用局部特征来描述目标。在 co-training 学习框架下,第一个分类器使用 Haar 特征。Haar 特征的优势是计算简单有效,基于 Haar 特征的 AdaBoost 算法在人脸检测中已经得到了成功的应用。局部二值模式特征(local binary pattern, LBP)也是一种很好的选择,该特征对光照变化影响更加鲁棒,在人脸识别中得到了很好的运用,但当样本很少时,会丢失重要的统计特性。本文的跟踪分类问题正是在小样本下执行的,因此, LBP 特征不适合目标跟踪分类问题。第二个分类器选择的是局部梯度直方图特征(histogram of oriented gradients, HOG),该特征值基于 sift 算法理论,已经在行人检测等领域有了成功的应用。

2 基于特征选择的 on-line Boosting 算法

与 off-line Boosting 一样, on-line Boosting 通过选择弱分类器 $\{h_n(x) | h_n(x): \chi \rightarrow \{1, -1\}\}$ 联合成一个强分类器 $H(x)$, $H(x) = \text{sign}(f_n(x))$ 。其中 $f_n(x) = \sum_{n=1}^N \alpha_n h_n(x)$, 其目标是最小化训练错误率。

Off-line Boosting 和 on-line Boosting 最大的不同在于数据的输入方式和强分类器 $H(x)$ 的构建方式。在 off-line 训练中,所有的数据是事先得到的,一次性输入到模型中去训练;而在 on-line 训练中,每一时刻只能得到一个数据,因此只能依次输入到模型中去训练。对于强分类器 $H(x)$,在 off-line 训练中是通过每一轮累加弱分类器得到的;而在 on-line 训练中是事先规定好弱分类器个数,每轮再进行更新。

这里对 Grabner 等人^[3] 的基于特征选择的 on-line Boosting 算法作简单介绍。该算法的主要思想基于特征选择器(selector)的 on-line Boosting,而不是弱分类器的直接更新。每一个 selector 在 M 个弱分类器 $H^n = \{h_1^n(x), \dots, h_M^n(x)\}$ 中进行选择。在训练过程中,selector 选择具有最小训练错误率的弱分类器来更新强分类器中的 h 。为了传递每个弱分类器对当前样本的分类情况,困难度因子 λ 贯穿于 N 个 selector 中。首先,selector 在集合 H^n 中被随机初始化,当一个新的样本 (x, y) , $y \in \{1, -1\}$ 到达时, N 个 selector 被顺序更新。权重 λ 的初始值为 1,具有最小训练错误率的弱分类器被 selector 选择。然后,相关的投票权重 α_n 和困难度因子 λ 被更新,执行到下一个 H^{n+1} 。如果样本被当前的 selector 错分, λ 增加;否则减少。最后,强分类器是由 N 个被各自 selector 选择出的弱分类器线性加权组成。

3 强分类器的置信度分析

在协同训练过程中,对每个强分类器的输出值进行精确

的置信度估计是提高目标跟踪准确率的重要因素。本文参照文献[11]的方法。已知训练集 $\chi^l = \{(x_1, y_1), \dots, (x_{1 \times l_1}, y_{1 \times l_1}) | x_i \in R^m, y_i \in \{-1, +1\}\}$, Boosting 在每个阶段都最小化指数损失函数 $L = \sum_{x \in \chi^l} e^{-yH(x)}$, 可被看做是一个可加逻辑回归问题(additive logistic regression), 因此,置信度测量式可被表示为 $P(y=1|x) = \frac{e^{H(x)}}{e^{H(x)} + e^{-H(x)}}$ 。

4 实验

本文采用 PETS2001 和 NUDT 序列图像作为测试集, 实验环境: CPU 为 Intel(R) Pentium(R) 处理器 1.70 GHz; 内存为 2 GB。在实验中,前 5 帧由人工标定, 实验结果如表 1 所示。

表 1 实验结果

测试集	测试帧数	文献[5]方法	文献[9]方法	本文的方法
PETS2001	453	386	423	451
NUDT	321	281	301	316

表 1 分别表达了三种不同方法的目标跟踪识别率。图 3 是在 PETS2001 实验中,第 200、253、330、397、480、592 帧目标跟踪结果,第一列标出了目标跟踪的位置,第二列表达了到当前帧及以前帧的跟踪目标处的最大置信度轨迹。从实验结果不难看出,与文献[5,9]相比,本文的方法能够更稳定地跟踪目标,跟丢的帧数最少,这主要得益于将 co-training 学习框架和基于特征选择的 on-line Boosting 有效结合起来。Co-training 学习框架能够尽可能地减少输入新样本带来的误差,而基于特征选择的 on-line Boosting 不仅训练速度快,采取的局部特征与文献[9]相比,更有利于目标跟踪。

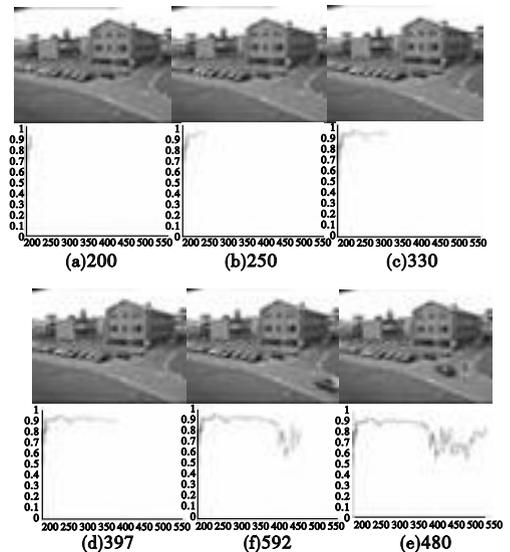


图 3 PETS2001 实验结果

5 结束语

针对视频目标跟踪问题,本文提出了一种基于 co-training 框架下的在线学习跟踪方法,实验证明了该方法的有效性。在未来的工作中,将重点研究基于 co-training 框架下的有遮挡的目标跟踪问题。

参考文献:

[1] AVIDAN S. Ensemble tracking [C]// Proc of CVPR. 2005: 494-501.
 [2] LEPETIT V, LAGGER P, FUA P. Randomized trees for real-time keypoint recognition [C]// Proc of CVPR. 2005: 775-781.

选取的边缘检测方法的检测结果,分析两种方法的边缘定位精确度和抗噪声干扰能力,如图 3 所示。实验中的两种边缘检测算法除阈值选取方法不同外,其他检测步骤完全一致。不同的图像可根据人眼对景物的关注侧面不同而手工调整阈值选取比例(该比例可确定各边缘点局部区域内阈值的选取)。本实验三幅图像均设定阈值选取比例为 80%。

表1 两种阈值检测方法边缘检测结果及评价

实验序号	提取的边缘点数		本文候选边缘点阈值获取		两算法综合评价
	Canny 检测法	本文算法	改进前 /ms	改进后 /ms	
1	1562	2126	1516	328	Canny边缘检测法存在丢边现象,本文算法准确定位边缘
2	1346	2127	1438	234	Canny边缘检测法受噪声影响较大,本文算法未受噪声影响
3	1793	1943	2891	609	Canny边缘检测法存在丢边现象,本文算法准确定位边缘且抑制噪声

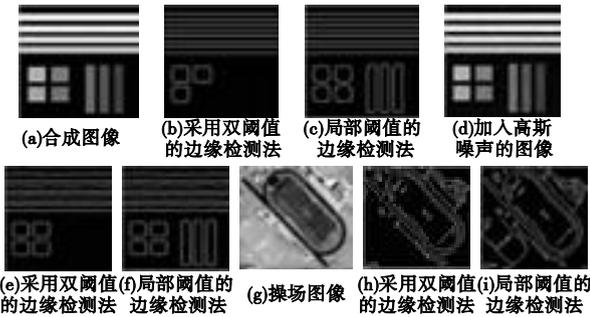


图3 采用不同阈值选取方法的边缘检测结果

实验 1 通过检测已知边缘的图 3(a),比较两阈值选取算法的边缘检测精确度。图 3(a)为 165 × 165 的 8 位灰度合成图像,各矩形区域内灰度相同,各区域间灰度值等差递减,且无噪声干扰,边缘点总数为 2 126。图 3(b)为采用全局阈值选取的 Canny 算法对图 3(a)边缘检测的结果,σ = 1(高斯函数标准差),高阈值 τ₂ 取全局梯度模值图像直方图个数比例 80% 处。图 3(c)为本文提出的基于局部区域动态阈值法的边缘检测算法作用于图 3(a)的结果,σ = 1, r = 30,局部高阈值 τ₂ 为局部区域梯度模值图像直方图个数比例 80% 处。将结果图 3(b)与(c)对比,全局阈值选取法存在丢边现象,本文提出的局部阈值选取法无丢边现象。表 1 记录了实验 1 中各方法的边缘点检测数量。

实验 2 比较两阈值选取算法的抗噪性。将图 3(a)加入信噪比为 0.005 的高斯噪声生成图 3(d)。图 3(e)为采用全局阈值选取的 Canny 算法对图 3(d)边缘检测结果,参数设置与实验 1 中一致。图 3(f)为本文提出的基于局部区域动态阈值提取的边缘检测算法作用于图 3(d)的结果,参数设置与实验 1 中一致。将结果图 3(e)与(f)对比观察,全局阈值选取法存在

丢边现象,本文提出的阈值选取法无丢边现象;将结果图 3(b)与(e)对比,(e)与(f)对比,全局阈值选取法的边缘检测结果易受噪声影响(同一比例的阈值选取方法中,噪声增加,则小梯度值的边缘点数增加,随之阈值减小,且边缘结果图中边缘点增加),本文提出的局部阈值选取法的检测结果基本无噪声影响。表 1 记录了实验 2 中各方法的边缘点检测数量。

实验 3 图 3(g)为 168 × 168 的 8 位灰度图像,通过检测真实机场图像(g)的边缘,综合对比阈值选取算法的边缘检测效果。图 3(h)为采用全局阈值选取的 Canny 算法对图 3(g)边缘检测结果,参数设置与实验 1 中一致;图 3(i)为本文提出的基于局部区域动态阈值提取的边缘检测算法作用于图(g)的结果, r = 30,其他参数设置与实验 1 中一致。实验结果图 3(h)与(i)对比,全局阈值选取法存在丢边现象,丢失了左下角操场边缘;本文提出的阈值选取法检测出了图中重要边缘。表 1 记录 3 实验 3 中各方法的边缘点检测数量。

5 结束语

本文针对现有的边缘检测方法中选取阈值不当,造成重要边缘丢失的问题,提出了基于局部区域的动态阈值选取方法,解决了现有局部区域阈值选取方法实用性差的问题。实验表明,本文采用的基于局部区域动态阈值选取的边缘检测方法,较全面地提取图像中各类边缘,不易受噪声影响;文中提出的局部区域快速阈值选取法有较低的时间复杂度。

参考文献:

[1] CANNY J. A computational approach to edge detection [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1986, 8(6): 679-698.

[2] ELMABROUK A, AGGOUN A. Edge detection using local histogram analysis. [J] Electronics Letters, 1998, 34(12): 1216-1217.

[3] WAN Shuai, YANG Fu-zheng, HE Ming-yi. Gradient-threshold edge detection based on perceptually adaptive threshold selection[C]//Proc of the 3rd Conference on Industrial Electronics and Applications. 2008:999-1002.

[4] ROSIN P, WEST G A W. Segmentation of edges into lines and arcs [J]. Image and Vision Computing, 1989, 7(2): 109-114.

[5] VOORHEES H, POGGIO T. Detecting textons and texture boundaries in natural images [C]//Proc of the 1st International Conference on Computer Vision. 1987:250-258.

[6] 罗宏文, 马驹良, 徐中宇. 基于统计的显微图像边缘检测阈值分析方法[J]. 吉林大学学报:理学版, 2006, 44(3): 397-400.

[7] 华刚, 郑南宁, 薛建儒. 基于改进遗传算法的边缘检测阈值自动选取及其应用[J]. 小型微型计算机系统, 2002, 23(3): 318-321.

[8] YU Yuan-hui, CHANG C C. A genetic-base adaptive threshold selection method for dynamic path tree structured vector quantization [J]. Image and Vision Computing, 2005, 23(6): 597-609.

(上接第 771 页)

[3] GRABNER H, BISCHOF H. On-line Boosting and vision [C]//Proc of CVPR. 2006: 260-267.

[4] GRABNER M, GRABNER H, BISCHOF H. Learning features for tracking [C]//Proc of CVPR. 2007:1-8.

[5] GRABNER H, SOCHMAN J, BISCHOF H, et al. Training sequential on-line Boosting classifier for visual tracking [C]//Proc of ICPR. 2008: 1-4.

[6] BLUM A, MITCHELL T. Combining labeled and unlabeled data with co-training [C]//Proc of the 11th Annual Conference on Computational Learning Theory. New York: ACM Press, 1998: 92-100.

[7] COLLINS M, SINGER Y. Unsupervised models for named entity clas-

sification [C]//Proc of Empirical Methods in Natural Language Processing. 1999.

[8] JAVED O, ALI S, SHAH M. On-line detection and classification of moving objects using progressively improving detectors [C]// Proc of CVPR. 2005: 695-700.

[9] 王路, 卓晴, 王文渊. 基于 Co-training 的协同目标跟踪 [J]. 计算机工程, 2009, 35(3): 202-204.

[10] OZA N C. On-line ensemble learning [D]. Berkeley: University of California, 2002.

[11] FRIEDMAN J, HASTIE T, TIBSHIRANI R. Additive logistic regression: a statistical view of Boosting [J]. Annals of Statistics, 2000, 28(2): 337-407.