

在线学习多重检测的可视对象跟踪方法

权 伟¹, 陈锦雄¹, 余南阳²

(1. 西南交通大学信息科学与技术学院, 四川成都 610031; 2. 西南交通大学机械工程学院, 四川成都 610031)

摘 要: 为了研究无约束环境下长时间可视跟踪问题, 提出了一种在线学习多重检测的对象跟踪方法. 该方法以随机蕨作为基础检测器结构, 通过在线学习的方式, 将目标对象的整体和局部表现, 以及由场景学习中发掘的同步对象同时作为检测学习的基础数据, 该检测器因而具备了对这多种对象的独立检测能力. 由于其各个检测部分发挥了各自不同的作用, 本文从测量的角度将检测器对这三种对象检测的结果进行融合, 通过计算检测关于目标的配置概率进而确定目标位置, 实现对象跟踪任务. 基于真实视频序列的实验结果验证了本文方法的有效性和稳定性, 以及较现有的跟踪方法在跟踪性能上的提高.

关键词: 对象跟踪; 多重检测; 在线学习; 随机蕨

中图分类号: TP319

文献标识码: A

文章编号: 0372-2112 (2014)05-0875-08

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

DOI: 10.3969/j.issn.0372-2112.2014.05.007

Online Learning of Multiple Detectors for Visual Object Tracking

QUAN Wei¹, CHEN Jin-xiong¹, YU Nan-yang²

(1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu, Sichuan 610031, China;

2. School of Mechanical Engineering, Southwest Jiaotong University, Chengdu, Sichuan 610031, China)

Abstract: In order to study the problem of long-term visual tracking in unconstrained environments, this paper proposes a method of learning multiple detectors online for visual object tracking. The method uses the random ferns as the basic detector. The entire and the local appearances of the target and the connected objects which are explored by the context learning are used synchronously as the training data to build and upgrade the object detector on the fly. Thus it is able to detect the objects with different classes independently. Since different detections are related to different object classes, the results of object detections are fused as the measurements and the probabilities of configuration hypotheses for the measurements to the target are calculated to find the target location for visual tracking task. Experimental results based on the real-world video sequences validate the effectiveness and robustness of our approach and demonstrate its better tracking performance than several state-of-the-art methods.

Key words: object tracking; multiple detectors; online learning; random ferns

1 引言

对象跟踪作为智能视频分析的关键问题, 在计算机视觉领域具有广泛的应用, 如智能监控、人机交互、机器人技术以及多媒体应用等. 尽管对此研究人员做出了大量的工作, 但是由于真实世界的复杂性, 如背景干扰、表现和光照变化、低图像分辨率以及帧跳跃等问题, 使得在无约束环境下实现对未知目标进行长时间实时稳定的可视跟踪任务仍然是一项极具挑战性的课题. 目前, 基于机器学习特别是在线学习的跟踪方法成为该领域研究的热点. 这是由于与传统的方法^[1~7]相比, 基于在

线学习的跟踪方法本质上更能够适应对象及其环境的变化, 有利于完成长时间的跟踪任务.

跟踪过程中在线学习的目的在于发掘未知的数据结构, 对此已逐步发展出一系列自适应的对象跟踪方法. 文献^[8~10]采用自学习的方式, 用接近和远离目标的样例更新对象模型, 然而, 预测目标一旦出错, 跟踪则无法继续. 为了克服这个问题, Yu 等^[11]提出了通过协作训练获得一个可再生的判别分类器, 从而实现重检测和失败恢复. 因此, 对象跟踪或者检测也被看作是一个分类问题, 即通过训练得到的分类器来判断该区域是目标还是背景. 由于跟踪的实时性要求, 相比于支持向量

机^[12]和主成分分析的方法^[13,14],基于 Boosting 的方法^[15~17]和基于随机森林的方法^[18~24]是目前主要研究的跟踪方法.而以上这些方法需要共同面对的一个难题是,如何既保证跟踪系统对目标变化的适应能力,又保证学习的准确性,避免漂移而导致跟踪失败.为此, Babenko 等^[25]提出了袋学习的思想和 OMB 算法(Online-MILBoost),在很大程度上提高了跟踪性能. Kalal 等^[26]将自适应跟踪与对象检测相结合,提出了 TLD (Tracking-Learning-Detection) 跟踪学习的框架,用以在跟踪过程中进行有约束的学习,即所谓的增长和裁剪学习策略,其跟踪系统 PNT^[27]采用 LK^[1](Lucas-Kanade Tracker)跟踪和随机森林检测^[28],取得了很好的实时跟踪效果.然而 PNT 容易在目标初始就快速运动的情况下(包括摄像头剧烈运动)跟踪失败,同时难以跟踪非刚性对象(如行人)以及处理快速的表现变化.对此,霍夫森林^[29~32]提供了一种可能的解决方法.霍夫森林是一种融合霍夫变换^[33]的随机森林^[18],它由多个决策树组成,每棵树将图像或者视频中的局部表现映射到它的叶节点中,而每个叶节点则包含在霍夫空间中的概率投票.由此,对象的定位或者检测被看作是寻求在霍夫图像中的概率极大值点.然而这样的检测过程十分耗时,因此基于霍夫森林的对象跟踪方法^[32]在实时性上远不及基于随机森林的方法^[26,27].另一方面,场景越复杂,跟踪或者检测就变得越困难,由于计算资源的局限和效率的要求,对象表现模型不能过于复杂.实际上,除了目标本身以外,背景信息也是十分有用且重要的信息. Yang 等^[34]采用图像分割融合方法,通过在运行时分析并获取辅助对象作为跟踪协助,对一定的跟踪场景取得了较稳定的跟踪性能,然而这种依赖辅助对象进行运动假设的跟踪难以适应更复杂的跟踪环境. Grabner 等^[35]提出通过获取目标周围有价值的特征点,借此预测目标的位置,增强了跟踪的稳定性,然而检测和匹配所有这些局部特征点的方法在计算上十分耗时. Thang 等^[36]通过将 PNT^[27]作为基础跟踪和同时增加对误匹配项的跟踪来改进 Grabner 等^[35]的算法,能够在一定程度上区分与目标相似的对象. Fan 等^[37]提出学习具有较强区别性的关注区域用以辅助跟踪,然而在场景出现剧烈运动时,这些局部区域限制和降低了其跟踪性能. Godec 等^[38]通过对场景进行聚类学习,将背景分类为多个虚拟的类型,取得了较为满意的跟踪效果,但是该方法假设背景只是逐步而细微的改变,这在很多跟踪场合并不成立,因此其应用有限.

鉴于此,本文提出一种在线学习多重检测的可视对象跟踪方法.该方法以随机森林作为基础检测器结构,通过在线学习的方式同步学习目标对象的整体和局部表现,以及由场景学习中发掘的同步对象,使得其检测

器具备了对这多种对象的独立检测能力.这种多重检测的各个检测部分将发挥各自不同的作用:在正常跟踪过程中,目标整体检测配合短时跟踪即可实现稳定的跟踪;当目标被部分遮挡时,目标局部检测可发挥主要作用;而当目标消失时(被完全遮挡或者移出视场),同步对象检测可作为预测目标的主要数据.本文从测量的角度将检测器对这三种对象检测的结果进行融合,通过计算检测关于目标的配置概率进而确定目标位置,实现对象跟踪任务.本文的创新点为:(1)提出了运行时对多种对象进行同步检测和学习的思想和方法;(2)提出了实时发掘场景中同步对象的方法;(3)提出了通过融合检测数据和最大化配置概率定位目标位置的计算方法.基于真实视频序列的实验结果验证了本文方法的有效性和稳定性,以及较现有的跟踪方法在跟踪性能上的提高.

2 学习多重检测的跟踪方法

2.1 多重检测器及其学习

多重检测指检测器具备多样检测能力,即用于检测器学习的数据类型呈多样性,因此能够检测多种类型的对象.不同的对象类型将对应不同的学习数据,检测器将通过跟踪过程中在线学习这些数据,不断提升其对象检测和识别能力.本文多重检测器基于具有实时检测能力的随机森林^[28]构建.用于检测器学习的样例均为图像块,并按类型区分,因此检测器能够对各种类型的对象进行独立的检测.

由于对比像素点差异适合于小块矩形图像的分析,因此本文采用像素点比较作为图像特征.该特征可计算任意两个不同位置的像素亮度值的差,如果该差值大于 0,则特征值为 1;否则为 0.如图 1 所示,组成一个像素点比较特征的两个像素点用相同的颜色表示.

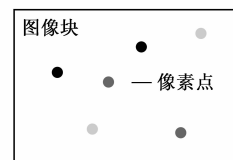
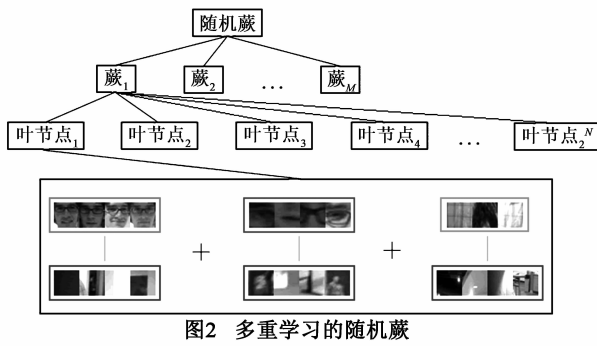


图1 像素点比较特征

随机森林由多个森林组成,每个森林包含多个叶节点,如图 2 所示.对检测器的学习即在相应的叶节点中增加对应类型包含的样例数目.设随机森林检测器包含 M 个森林,每个森林包含 N 个不同的二元特征(这里为像素点比较特征),则每个森林包含 2^N 个叶节点,每个叶节点对应一个 N 位的二进制编码值,即特征编码值.将样例数据逐个输入检测器,每个森林的叶节点将根据样例数据的二进制编码值,记录落在该节点上的样例类型及其对应的数量(初始时为 0).



本文检测器的学习数据包括目标对象的整体表观和局部表观,场景对象以及各自对应的背景图像.因此,检测器的每个叶节点中将记录共6种类型的图像块数量,如图2所示.

对于整体表观的学习,根据目标所在的区域提取图像块作为正样例,其类型表示为 c_T^+ ;在远离目标区域所提取的图像块作为负样例,其类型表示为 c_T^- ;正、负样例图像块的大小与目标相同.在跟踪初始,目标对象可通过手动选择或者自动检测得到;而在跟踪过程中,目标对象则由跟踪系统确定其位置和大小.如图3所示,红色矩形框表示选取的正样例图像块,蓝色矩形框表示选取的负样例图像块(部分),即背景图像块,其中,图3(a)表示跟踪初始目标选择后的样例选取情况,图3(b)表示跟踪过程中的样例选取情况.最后将这些样例图像均规则化为一个统一的尺寸大小,如 25×25 像素大小,并作为学习样例输入检测器.

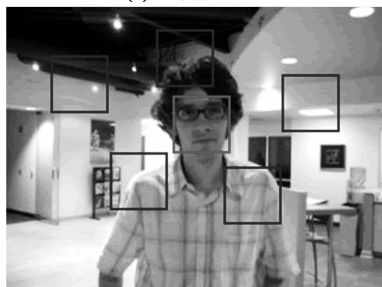
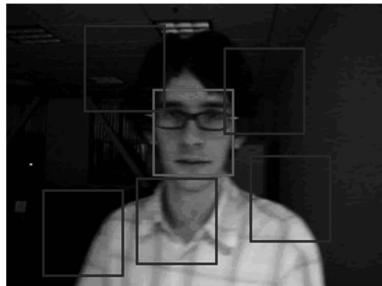


图3 选取整体表观样例图像

局部表观的学习过程与整体表观的学习过程相同,仅学习的对象数据不同.局部表观学习的正样例为

目标的局部图像,其类型表示为 c_L^+ ;而在目标区域外部提取的图像块作为负样例,其类型表示为 c_L^- .这里作为负样例的背景图像块尺寸小于整体表观学习中的背景图像块.正样例图像块可在目标对象的每个位置以不同的尺寸提取局部图像得到,如图4所示,紫色矩形框为在目标对象内部为局部检测器提取的正样例,蓝色矩形框为在背景提取的负样例.最后将这些样例图像规则化后(如 15×15 像素大小)作为学习数据输入检测器.



图4 选取局部表观样例图像

场景对象的学习旨在发掘场景中存在的同步对象.同步对象指与目标具有相同运动结构的对象,如与目标具有相同的运动方向,速度以及加速度,或者与目标对象的距离始终保持不变.由于同步对象在跟踪过程中是未知的,因此需要在运行时分析和发掘这样的对象.对此,首先将整个图像划分为 $N \times M$ 个子图像区域,如图5所示,根据图像像素的亮度值计算每个子图像区域的均值和标准差.采用 K-means 聚类算法对这些子区域进行聚类分析,如果某个聚类的标准差大于阈值 θ_σ ,则将其分裂为两个聚类类型.然后按照逐行扫描方法将具有相同类型的子区域合并为连通区域.具体为,如果某个子区域的上、下、左、右的四个领域内存在与其具有相同类型的子区域,则将它们合并到一起.接着对合并后的连通区域提取最大矩形区并作为待选对象,该矩形区为对应连通区域外只包含相同类型子区域的最大矩形范围,接着删除那些过大或者过小的连通区.如图5所示,绿色矩形框表示候选同步对象.因此,每个这样的候选同步对象将在一定时间间隔的图

像序列中多次出现或者在某帧后消失. 如果在这段时间内某个候选同步对象与目标同时存在, 且与其具有相同的运动, 则将该候选同步对象作为目标真正的同步对象. 如图 5 所示, 紫色矩形框表示目标的同步对象. 由此, 将得到的同步对象作为学习的正样例, 其类型表示为 c_D^+ ; 而在其余背景中选择的图像块作为负样例 (图 5 中由蓝色矩形框表示, 其大小与对应的正样例相同), 其类型表示为 c_D^- . 如果同步对象不只一个, 则需在检测器中通过类型和序号标记对它们加以区分. 最后将这些样例图像规则化后 (如 25×25 像素大小) 作为学习数据输入检测器.



图5 场景对象学习

2.2 对象检测与跟踪

对象检测的根本作用在于, 当短时跟踪失败后, 可在目标重新出现时通过检测定位来恢复对目标的跟踪. 本文多重检测器将同时对整个目标对象, 目标对象的局部, 以及场景中的同步对象进行检测.

本文采用搜索窗方法实现对象检测, 即对图像的每一个位置以多个不同的尺度选取矩形框图像, 规则化后作为测试样例输入检测器. 然后计算测试样例属于类型 c 的概率 p_c , 其计算公式为:

$$p_c = \frac{1}{M} \sum_{m=1}^M p_c^m,$$

其中, $p_c^m = \frac{N_c}{N_s}$ 表示第 m 个蕨判断测试样例属于类型 c 的概率, $c \in \{c_T^+, c_L^+, c_D^+\}$; M 为检测器中包含的蕨的数量; N_c 表示测试样例落入第 m 个蕨对应的叶节点中属于类型 c 的样例数量, N_s 表示该节点中类型 c 对应的正、负样例的总数. 因此, 对于整个目标对象, $N_s = N_{c_T^+} + N_{c_T^-}$; 相应地, 对于目标局部对象 $N_s = N_{c_L^+} + N_{c_L^-}$, 对于场景同步对象 $N_s = N_{c_D^+} + N_{c_D^-}$.

如果 p_c 大于阈值 θ_0 , 则认为该测试样例为类型 c 的对象. 因此, 对于一帧图像, 检测器返回的检测结果中将包含多种类型的多个对象. 根据检测结果, 则可以有多多个不同的目标配置, 目标配置即根据对象检测结果设置其所在的位置和大小. 设 Z_T 为对整个目标对象的检测结果, Z_L 为对目标对象局部的检测结果, Z_D 为

对场景同步对象检测的结果. Ω_i 表示第 i 种目标配置, 则该配置的概率 p_{Ω_i} 计算为:

$$p_{\Omega_i} = p(\Omega_i | Z_T, Z_L, Z_D). \text{ 根据贝叶斯公}$$

式有:

$$p_{\Omega_i} = p(\Omega_i | Z_T, Z_L, Z_D)$$

$$= p(Z_T, Z_L, Z_D | \Omega_i) p(\Omega_i) / p(Z_T, Z_L, Z_D).$$

因此, 跟踪任务即在图像中找到使 p_{Ω_i} 最大的配置 Ω^* , 即:

$$\Omega^* = \operatorname{argmax}_{\Omega_i \in G} p_{\Omega_i} = \operatorname{argmax}_{\Omega_i \in G} (p(Z_T, Z_L, Z_D | \Omega_i) p(\Omega_i) / p(Z_T, Z_L, Z_D)),$$

其中, G 表示所有配置的集合. 由于检测器对各种类型的对象进行独立检测, 因此可以认为 Z_T, Z_L, Z_D, Ω_i 相互独立, 且 $p(\Omega_i), p(Z_T, Z_L, Z_D)$ 为常数, 则有:

$$\begin{aligned} \Omega^* &= \operatorname{argmax}_{\Omega_i \in G} (p(Z_T, Z_L, Z_D | \Omega_i) p(\Omega_i) / p(Z_T, Z_L, Z_D)) \\ &= \operatorname{argmax}_{\Omega_i \in G} p(Z_T, Z_L, Z_D | \Omega_i) = \operatorname{argmax}_{\Omega_i \in G} (p(Z_T | \Omega_i) p(Z_L | \Omega_i) p(Z_D | \Omega_i)) \end{aligned}$$

为了得到 Ω^* , 接下来对以上公式中的 $p(Z_T | \Omega_i), p(Z_L | \Omega_i), p(Z_D | \Omega_i)$ 分别进行计算. 其中 $p(Z_T | \Omega_i)$ 可计算为:

$$p(Z_T | \Omega_i) = q_d q_h,$$

其中, q_d 为检测器根据整体目标对象的检测概率, 即 $q_d = p_{c_T^+}$; q_h 为目标运动的预测概率, 计算为:

$$q_h = \frac{1}{\sqrt{2\pi}\sigma_T} \exp\left(-\frac{\|x_i^t - x_o^{t-1}\|^2}{2\sigma_T^2}\right),$$

其中, x_i^t 为 t 时刻检测目标的位置, x_o^{t-1} 为 $t-1$ 时刻跟踪系统已确定的目标位置; 标准差 σ_T 可根据具体情况设置, 如 $\sigma_T = 20$. 而 $p(Z_L | \Omega_i)$ 可通过统计目标所在区域中检测到的局部对象的数目及其概率得到, 如图 8 所示, 绿色矩形框表示目标, 也表示用于统计局部对象数目的区域. 因此 $p(Z_L | \Omega_i)$ 可计算为:

$$p(Z_L | \Omega_i) = \sum_{i=1}^N q_i / (W_T \times H_T),$$

其中, W_T, H_T 分别为目标矩形的宽度和高度; q_i 为第 i 个局部对象的检测概率, 即 $q_i = p_{c_L^+}$. 最后, $p(Z_D | \Omega_i)$ 可根据同步对象预测目标的位置与检测位置之间的距离来计算, 即:

$$p(Z_D | \Omega_i) = \frac{1}{\sqrt{2\pi}\sigma_D} \exp\left(-\frac{\|x_i - x_D\|^2}{2\sigma_D^2}\right),$$

其中, x_i 为目标检测位置, x_D 为同步对象预测的目标位置; 同理, 标准差 σ_D 可根据具体情况设置, 如 $\sigma_D = 20$.

3 实验及结果

本文的短时跟踪采用 NCC-C(Normalized Cross-Corre-

lation and Color)方法,这是由于 NCC-C 具有较高的跟踪精度和稳定性,具体请参见文献[39].用于短时跟踪相似度计算的在线模型所包含的图像块规则化为 25×25 像素大小,表示对象与在线模型接近程度的阈值为 0.8.如果跟踪过程中已确定的目标与在线模型的相似度小于 0.5,则将当前目标表观加入到在线模型.为保证合理的运行时间,短时跟踪搜索区域为 50×50 .在本文的实现中,检测器包含 8 个蕨,每个蕨包含 15 个像素对比较特征;用于检测器训练和学习的样例图像块均规则化为 15×15 像素大小;检测阈值 θ_0 为 0.5.对于 320×240 大小的图像,将其划分为 21×16 个子图像区域(即每个子图像区域 15×15 像素大小)用于场景聚类学习,其聚类标准差阈值 θ_c 按照实验经验结果设置为 30.获取同步对象的时间间隔为 10 帧.配置概率计算中 $\sigma_T = \sigma_D = 20$.本文基于 C# 多线程编程实现的跟踪系统,在一个普通双核 PC 机上的平均运行速率为 3fps.实验所用的视频序列来自目前国际上被广泛使用的基准测试序列 Babenko 序列[25].

为了说明各个检测部分对跟踪性能的影响,本文将三种检测及其组合进行跟踪测试.评价指标为跟踪稳定性和跟踪精度.跟踪稳定性为关于整个视频序列目标被正确跟踪的百分比.具体为,如果某帧中跟踪矩形框有 30% 没有包含目标或者目标有 30% 没有被跟踪矩形框包含,则认为该帧跟踪错误;否则跟踪正确.跟踪精度为跟踪矩形框与目标的平均中心位置误差,计算为: $1 - \text{平均中心位置误差}/H$, 其中 H 为最大误差宽度,这里 $H = 30$.如图 6 所示为 David 序列下基于各种检测组合的跟踪量化结果,其中,ED 表示整体对象检测,LD 表示局部对象检测,CD 表示场景对象检测;它们的组合即表示同时进行其多种检测,分别表示为 ED + LD, ED + CD, LD + CD, 以及 ED + LD + CD.从图 6 可以看出,LD 略为优于 ED,而 CD 在没有 ED 和 LD 配合的情况下最差;ED + LD 的组合优于其它两种检测的组合;最好的结果为 ED + LD + CD,这也是本文采用的方法.图 7 为 David 序列下本文方法的跟踪量化结果,其中重

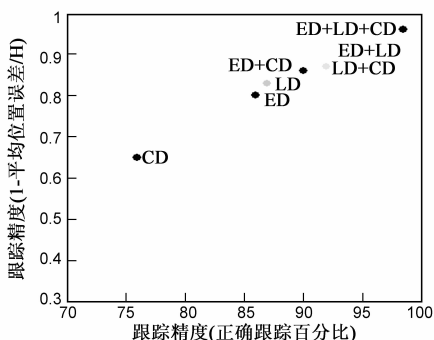


图6 各种检测方法关于David序列的跟踪性能

叠率按照文献[40]的方法计算,可以看出本文方法在跟踪过程中保持了较高的跟踪稳定性.

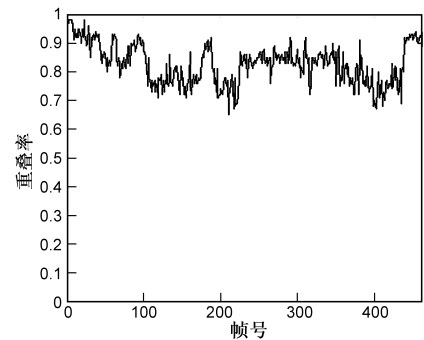


图7 本文方法关于David序列的跟踪重叠率

因此,本文多重检测方法中的各个检测部分发挥了各自不同的作用:在正常跟踪过程中,目标整体检测配合短时跟踪即可实现稳定的跟踪;当目标被部分遮挡时,目标局部检测可发挥主要作用;而当目标消失时(被完全遮挡或者移出视场),同步对象检测可作为预测目标的主要数据.对这些检测数据进行融合即计算乘积概率 $p(Z_T | \Omega_i)p(Z_L | \Omega_i)p(Z_D | \Omega_i)$.具体为,首先对当前帧图像进行目标整体检测、目标局部检测和同步对象检测.如果整体检测返回的检测结果中有属于目标的位置(检测概率 > 0.5),则按照前面介绍的计算方法对每个这样的候选目标位置计算其 $p(Z_T | \Omega_i)$,同时计算该候选目标区域内的 $p(Z_L | \Omega_i)$;如果存在同步对象,则对每个候选目标位置计算 $p(Z_D | \Omega_i)$,有多个同步对象则累加.由此计算它们的乘积 $p(Z_T | \Omega_i)p(Z_L | \Omega_i)p(Z_D | \Omega_i)$.如果整体检测返回的检测结果中没有属于目标的位置,则以上乘积中不考虑 $p(Z_T | \Omega_i)$,此时对整个图像计算其每个位置的 $p(Z_L | \Omega_i)$ 和 $p(Z_D | \Omega_i)$;同理,如果同步对象不存在,则以上乘积中不考虑 $p(Z_D | \Omega_i)$.最后该乘积最大值对应的候选位置为当前跟踪的目标位置.如图 8 所示为几种困难情况下的可视跟踪结果,其中绿色矩形框为跟踪目标,绿色虚线框为此刻整体检测检测到的另一个可能位置,紫色小框为局部检测结果,这里仅显示了整体检测框中部分具有较大概率的局部对象,而洋红色为场景中的同步对象.图 8(a)为 David 序列下当目标旋转时的跟踪情况,图 8(b)为 Face 1 序列下当目标被部分遮挡时的跟踪情况,可以看出由于局部检测的结果加上同步对象的预测,跟踪系统正确的跟踪到了目标.图 8(c)为 Girl 序列下当场景出现与目标相似对象时的跟踪情况,跟踪系统通过计算两个相似的整体检测中关于局部对象的统计概率实现了正确的目标跟踪,即右边“女孩”的脸具有更大的配置(测量)概率.因此本文方法具有一定的相似目标区分能力,如果加入对相似目标的同步跟踪,则

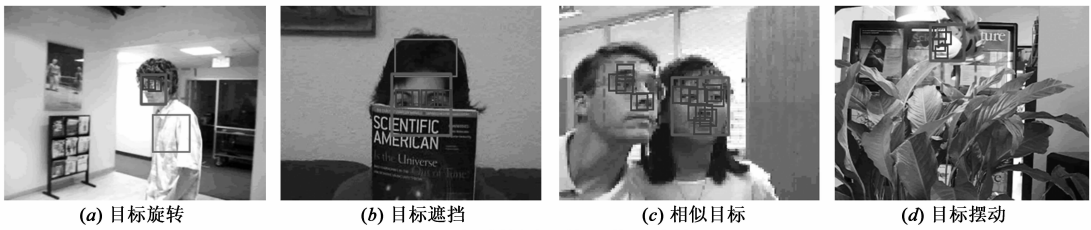


图8 多重检测下的几种跟踪情况

可以更好的对场景中出现的相似对象加以区分.图8(d)为 Tiger 1 序列下当目标摆动时的跟踪情况,跟踪系统没有可用的同步对象,这时通过整体检测和局部检测实现目标跟踪.

表1 Babenko 序列下各方法跟踪实验结果(正确跟踪百分比)

Video Sequence	Frames	OAB	ORF	FT	MILT	CDCT	PNT	CT	Ours
Coke	292	18.4	17.0	11.6	44.5	43.2	41.3	44.5	52.6
David	462	30.8	87.2	73.8	82.7	79.5	90.5	92.2	98.5
Face 1	888	72.9	97.5	97.5	89.6	93.7	95.6	96.8	99.7
Face 2	820	76.3	68.9	63.7	92.9	93.3	94.9	94.9	98.9
Girl	502	45.8	92.6	87.4	85.2	91.5	94.3	95.6	97.2
Sylvester	1345	56.1	69.2	82.6	82.4	83.1	85.7	87.9	93.0
Tiger 1	354	31.6	27.3	26.5	73.8	73.9	67.4	74.8	86.3
Tiger 2	365	35.5	20.8	23.5	76.5	46.5	71.8	83.7	89.8

采用以上的跟踪稳定度(即关于整个视频序列目标被正确跟踪的百分比),将本文的方法与其它的跟踪算法进行比较.用于比较的其它跟踪算法包括 OAB^[15]、ORF^[21]、FT^[41]、MILT^[25]、CDCT^[38]、PNT^[27]和 CT^[36].表1列出了 Babenko 序列下各跟踪方法的统计结果,其中最佳的跟踪结果用粗体表示,第二好的结果用带下划线的斜体表示.可以看出本文的方法总体上优于其它的跟踪算法.图9为本文方法在 Babenko 各序列下的可视跟踪结果(样例帧图像),其中绿色框表示跟踪的目标.

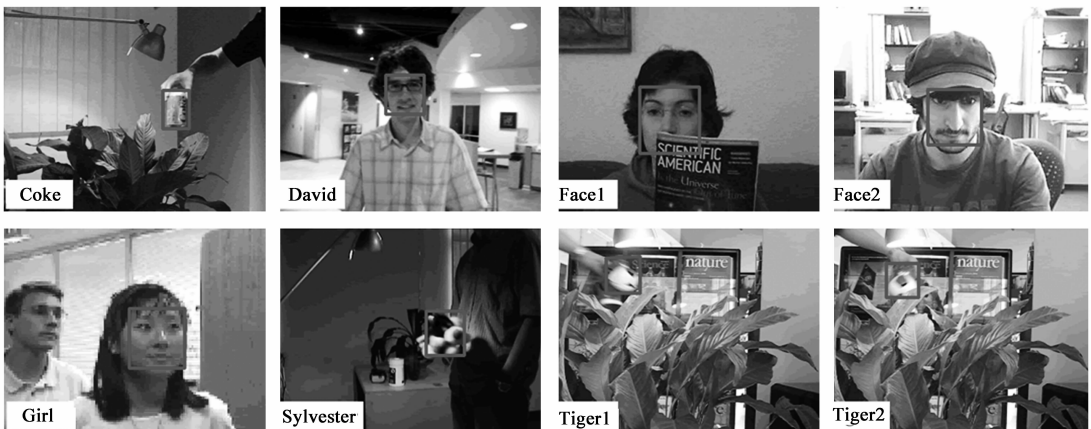


图9 Babenko序列下本文方法可视跟踪结果.框线表示跟踪目标

表2列出了各方法的跟踪效率(平均帧处理速率).由于本文的跟踪方法在跟踪过程中进行三种对象的检测和学习,包括发掘场景中的同步对象,并从测量的角度将检测器对这三种对象检测的结果进行融合,实现了跟踪性能上的提高.虽然目前我们基于 C# 实现的跟踪系统其运行速率相比其它方法略慢,但是它仍然很好的实现了实时的对象跟踪,尤其在跟踪的稳定性方面远优于其它的跟踪方法.因此,提高本文方法的算法运行效率是我们接下来需要进一步深入的研究工作.

表2 各方法跟踪效率(平均帧处理速率,单位:帧/秒)

OAB	ORF	FT	MILT	CDCT	PNT	CT	Ours
13	10	5	11	8	12	10	3

4 结论

本文提出了一种在线学习多重检测的可视对象跟踪方法,该方法通过同步学习目标对象的整体和局部表现,以及由场景学习中发掘的同步对象,实现对这多种对象的独立检测,其各个检测部分发挥了各自不同的作用.本文从测量的角度将检测器对这三种对象检测的结果进行融合,通过计算检测对目标的配置概率进而确定目标位置,实现对象跟踪任务.实验结果表明本文的方法能够处理更加复杂的跟踪环境,跟踪性能从总体上优于其它的跟踪算法.

今后我们将继续研究更加精练的在线学习方法,使得可以用更少的图像信息获得更加稳定快速的跟踪

性能.而另一个重要的问题则是怎样更好地选择学习样例,使这些样例图像块更具区别性.此外,如果能够更好地结合多核处理器系统实现或者在 GPU 上实现可以进一步减少计算时间,提高算法效率.

参考文献

- [1] B Lucas, T Kanade. An iterative image registration technique with an application to stereo vision[A]. Proceedings of the 7th International Joint Conference on Artificial Intelligence (IJ-CAI)[C]. Vancouver, Canada, 1981. 674 – 679.
- [2] M Isard, A Blake. CONDENSATION—Conditional density propagation for visual tracking [J]. International Journal of Computer Vision (IJCV), 1998, 29(1): 5 – 28.
- [3] 常发亮, 马丽, 刘增晓, 乔谊正. 复杂环境下基于自适应粒子滤波器的目标跟踪[J]. 电子学报, 2006, 34(12): 2150 – 2153.
CHANG Fa-liang, MA Li, LIU Zeng-xiao, QIAO Yi-zheng. Target tracking based on adaptive particle filter under complex background[J]. Acta Electronica Sinica, 2006, 34(12): 2150 – 2153. (in Chinese)
- [4] D Comaniciu, V Ramesh, P Meer. Kernel-based object tracking [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2003, 25(5): 564 – 5773.
- [5] 江淑红, 汪沁, 张建秋, 胡波. 基于目标中心距离加权和图像特征识别的跟踪算法[J]. 电子学报, 2006, 34(7): 1175 – 1180.
JIANG Shu-hong, WANG Qin, ZHANG Jian-qiu, HU Bo. An image tracking algorithm based on object center distance-weighting and image feature recognition [J]. Acta Electronica Sinica. 2006, 34(7): 1175 – 1180. (in Chinese)
- [6] 李培华. 一种新颖的基于颜色信息的粒子滤波器跟踪算法[J]. 计算机学报, 2009, 32(12): 2454 – 2463.
Li Pei-Hua. A novel color based particle filter algorithm for object tracking[J]. Chinese Journal of Computers, 2009, 32(12): 2454 – 2463. (in Chinese)
- [7] A Yilmaz, O Javed, M Shah. Object tracking: A survey [J]. ACM Computing Surveys, 2006, 38(4): 13.
- [8] S Avidan. Ensemble tracking [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2007, 29(2): 261 – 271.
- [9] R Collins, Y Liu, M Leordeanu. Online selection of discriminative tracking features [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2005, 27(10): 1631 – 1643.
- [10] J Lim, D Ross, R Lin, M Yang. Incremental learning for visual tracking [A]. Proceedings of Conference on Advances in Neural Information Processing Systems (NIPS) [C]. Vancouver, Canada: the MIT Press, 2004. 793 – 800.
- [11] Q Yu, T Dinh, G Medioni. Online tracking and reacquisition using co-trained generative and discriminative trackers [J]. Lecture Notes in Computer Science, 2008, 5303: 678 – 691.
- [12] S Avidan. Support vector tracking [J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2004, 26(8): 1064 – 1072.
- [13] DA Ross, J Lim, R Lin, M Yang. Incremental learning for robust visual tracking [J]. International Journal of Computer Vision (IJCV), 2008, 77(1-3): 125 – 141.
- [14] 温静, 李洁, 高新波. 基于增量张量子空间学习的自适应目标跟踪[J]. 电子学报, 2009, 37(7): 1618 – 1623.
WEN Jing, LI Jie, GAO Xin-bo. Adaptive object tracking with incremental tensor subspace learning [J]. Acta Electronica Sinica, 2009, 37(7): 1618 – 1623. (in Chinese)
- [15] H Grabner, H Bischof. On-line boosting and vision [A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR) [C]. New York, USA, 2006, 1: 260 – 267.
- [16] H Grabner, C Leistner, H Bischof. Semi-supervised on-line boosting for robust tracking [J]. Lecture Notes in Computer Science, 2008, 5302: 234 – 247.
- [17] S Stalder, H Grabner, L van Gool, E Zurich, K Leuven. Beyond semi-supervised tracking: tracking should be as simple as detection, but not simpler than recognition [A]. Proceedings of 12th IEEE International Conference on Computer Vision Workshops (ICCV Workshops) [C]. Kyoto, Japan, 2009. 1409 – 1416.
- [18] L Breiman. Random forests [J]. Machine Learning, 2001, 45(1): 5 – 32.
- [19] C Leistner, A Saffari, J Santner, H Bischof. Semi-supervised random forests [A]. Proceedings of 12th IEEE International Conference on Computer Vision (ICCV) [C]. Kyoto, Japan, 2009. 506 – 513.
- [20] C Leistner, A Saffari, H Bischof. MILForests: Multiple-Instance learning with randomized trees [A]. Proceedings of 11th European Conference on Computer Vision (ECCV) [C]. Crete, Greece, 2010. 29 – 42.
- [21] A Saffari, C Leistner, J Santner, M Godec, H Bischof. On-line random forests [A]. Proceedings of 12th IEEE International Conference on Computer Vision Workshops (ICCV Workshops) [C]. Kyoto, Japan, 2009. 1393 – 1400.
- [22] C Leistner, M Godec, A Saffari, H Bischof. On-line multi-view forests for tracking [J]. Lecture Notes in Computer Science, 2010, 6376: 493 – 502.
- [23] A Wang, G Wan, Z Cheng, S Li. An incremental extremely random forest classifier for online learning and tracking [A]. Proceedings of 16th IEEE International Conference on Image Processing (ICIP) [C]. Cairo, 2009. 1449 – 1452.
- [24] 王爱平, 万国伟, 程志全, 李思昆. 支持在线学习的增量式极端随机森林分类器 [J]. 软件学报, 2011, 22(9): 2059 – 2074.

- WANG Ai-Ping, WAN Guo-Wei, CHENG Zhi-Quan, LI Si-Kun. Incremental learning extremely random forest classifier for online learning[J]. Journal of Software, 2011, 22(9): 2059 – 2074. (in Chinese)
- [25] B Babenko, M-H. Yang, S Belongie. Visual tracking with on-line multiple instance learning[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Miami, 2009. 983 – 990.
- [26] Z Kalal, J Matas, K Mikolajczyk. Online learning of robust object detectors during unstable tracking[A]. Proceedings of 12th IEEE International Conference on Computer Vision Workshops (ICCV Workshops)[C]. Kyoto, Japan, 2009. 1417 – 1424.
- [27] Z Kalal, J Matas, K Mikolajczyk. P-N learning: bootstrapping binary classifiers by structural constraints[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. San Francisco, 2010. 49 – 56.
- [28] M Ozuysal, P Fua, V Lepetit. Fast keypoint recognition in ten lines of code[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Minneapolis, 2007. 1 – 8.
- [29] J Gall, V Lempitsky. Class-specific Hough forests for object detection[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Miami, 2009. 1022 – 1029.
- [30] J Gall, N Razavi, L Van Gool. On-line adaption of class-specific codebooks for instance tracking[A]. Proceedings of British Machine Vision Conference (BMVC)[C]. Aberystwyth, 2010. 1 – 12.
- [31] A Yao, J Gall, L Van Gool. A Hough transform-based voting framework for action recognition[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. San Francisco, 2010. 2061 – 2068.
- [32] J Gall, A Yao, N Razavi, L van Gool, Victor Lempitsky. Hough forests for object detection, tracking, and action recognition[J]. IEEE Trans Pattern Analysis and Machine Intelligence, 2011, 33(11): 2188 – 2202.
- [33] D H Ballard. Generalizing the Hough transform to detect arbitrary shapes[J]. Pattern Recognition, 1981, 13(2): 111 – 122.
- [34] M Yang, Y Wu, G Hua. Context-aware visual tracking[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2009, 31(7): 1195 – 1209.
- [35] H Grabner, J Matas, L V Gool, P Catin. Tracking the invisible: Learning where the object might be[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. San Francisco, 2010. 1285 – 1292.
- [36] T B Dinh, N Vo, G Medioni. Context tracker: Exploring supporters and distracters in unconstrained environments[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Providence, 2011. 1177 – 1184.
- [37] J L Fan, Y Wu, S Y Dai. Discriminative spatial attention for robust tracking[A]. Proceedings of 11th European Conference on Computer Vision (ECCV)[C]. Crete, Greece, 2010. 480 – 493.
- [38] M Godec, S Sternig, P M Roth, H Bischof. Context-driven clustering by multi-class classification in an active learning framework[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)[C]. San Francisco, 2010. 19 – 24.
- [39] B Stenger, T Woodley, R Cipolla. Learning to track with multiple observers[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. Miami, 2009. 1063 – 6919.
- [40] S Agarwal, A Awan, D Roth. Learning to detect objects in images via a sparse, part-based representation[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2004, 26(11): 1475 – 1490.
- [41] A Adam, E Rivlin, I Shimshoni. Robust fragments-based tracking using the integral histogram[A]. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)[C]. New York, USA, 2006. 798 – 805.

作者简介



权 伟(通信作者) 男, 1982 年出生于四川宜宾, 博士研究生。目前主要从事计算机视觉、模式识别、机器学习以及图像处理方面的研究。
E-mail: xueiquan@gmail.com



陈锦雄 男, 1962 年出生于四川成都, 博士。美国乔治梅森大学教授, 主要从事计算机图形、虚拟现实、计算机视觉以及数据可视化 and 仿真等方面的研究。