

基于 Co-Training 的协同目标跟踪

王 路, 卓 晴, 王文渊

(清华大学自动化系, 北京 100084)

摘 要: 运动目标跟踪是计算机视觉的核心问题之一, 广泛应用于诸多领域。该文提出一种基于 Co-Training 半监督学习框架的目标跟踪方法。该方法融合 2 种互相独立的特征信息来描述目标模型, 采用 Co-Training 来协同更新模型, 有效避免了现有方法的误差累积问题。实验结果证明, 该方法在复杂场景下仍能实现稳定有效的跟踪。

关键词: 目标跟踪; 联合训练; 半监督学习; 特征融合

Collaborative Object Tracking Based on Co-Training

WANG Lu, ZHUO Qing, WANG Wen-yuan

(Department of Automation, Tsinghua University, Beijing 100084)

【Abstract】 Moving object tracking is a key problem in computer vision and has many applications in various fields. This paper proposes a collaborative tracking method based on Co-Training frame work. The method fuses information from two types of features space to describe the object. The model is updated with Co-Training, which avoids the error accumulation problem. The experiment demonstrates the performance of the method under complex scenarios.

【Key words】 object tracking; Co-Training; semi-supervised learning; feature fusion

1 概述

运动目标跟踪是计算机视觉研究领域的一个重要课题, 在智能监控、机器人导航、视频内容分析等诸多领域广泛应用。所谓运动目标跟踪就是在视频中, 利用目标的有效特征, 寻找与目标模板最相似的图像区域, 从而获得目标的位置信息。近年来, 跟踪问题较多被分成目标和背景 2 类。文献[1]提出一种在线特征选择的方法, 提取最能判别目标和背景的颜色特征来区分出目标和背景。文献[2]采用集成学习的思想, 通过在线学习一组弱分类器的组合来区分出前景。文献[3]提出的跟踪方法, 核心思想就是基于增量空间方法的在线分类。由于在实际跟踪问题中, 目标和外界环境随着时间推移总会发生一些变化, 因此上述方法都采用了在线更新分类器的策略来适应这种变化。对于新获取的视频帧, 这些方法利用分类器区分出目标区域和背景区域, 并将这些区域作为新样本添加到分类器中去重新学习, 如此反复, 进而达到更新模型的目的。这一类利用分类器本身得到的样本来训练自身的过程在机器学习中称为“自学习”(self-training), 这种训练方法导致的直接结果就是分类误差的累积和分类器精度的下降^[4]。在跟踪过程中, 其后果尤为严重。

本文提出一种全新的目标跟踪框架, 该方法采用 Multi-View 的思想, 从 2 种独立的特征空间中抽取目标特征, 采用 2 个在线分类器协同工作来跟踪目标, 有效地解决了上述方法的自学习问题, 并得到更具鲁棒性的跟踪结果。

2 协同目标跟踪

2.1 Co-Training 与目标跟踪

Co-Training^[5]是半监督学习的一种。在实际的分类问题中, 对样本进行标记是一项代价昂贵的工程, 因此, 往往只有少部分样本会被标记, 而大量的样本则未被标记。这些未

标记的样本虽然不能直接用来训练分类器, 但可以从这些未标记样本中获取数据的一些信息来辅助训练分类器, 从而增强分类器的性能, 这类问题称作半监督学习。

对于半监督学习, 早期常用的方法就是自学习, 即首先利用标记的样本训练一个分类器, 然后用此分类器对未标记样本进行标记, 将置信度高的被标记样本与原有标记样本一起重新训练一个新分类器, 再不断重复这个过程。但是这种训练方法的一个缺陷就是稳定性不佳^[4], 很容易导致分类器的误差累积从而影响分类器性能。Blum 提出了 Co-Training 方法。这种方法基于 multi-view 思想, 即用 2 种互相独立的完备特征来描述对象, 对应地建立 2 个分类器, 它们将各自置信度较高的标记样本交与对方来更新训练分类器, 并不断重复这个过程。由于 Co-Training 融合了 2 类完备特征的信息, 在更新过程中巧妙地避开了“自学习过程”, 自推出以来, Co-Training 方法在文本分类等领域取得了较好的效果。

目标跟踪同样可以被视作一个半监督学习问题。由于视频数据的标记工作量巨大, 往往只有初始的几帧会被人工标记, 因此跟踪模型必须不断地从跟踪结果中学习目标和背景的变化信息。而所有上述基于分类的跟踪方法都首先利用已标记的样本来建立初始分类器, 在跟踪过程中, 利用新标记的样本来更新分类器, 因此, 这些方法都可以归结到自学习的框架中去。由于目标跟踪问题中, 目标和环境的不确定因素要比普通分类问题更加复杂, 由此带来的误差累积效益也更为严重, 使长时间的跟踪变得不稳定。基于跟踪问题和半监督学习的相似性, 合理地采用 Co-Training 框架, 可有效增

作者简介: 王 路(1980—), 男, 博士研究生, 主研方向: 计算机视觉, 模式识别; 卓 晴, 副教授; 王文渊, 教授、博士生导师
收稿日期: 2008-05-11 **E-mail:** l-wang02@mails.tsinghua.edu.cn

强目标跟踪算法的稳定性。

2.2 目标模型与 Two-view 特征提取

在具体案例中,目标模型的局部直方图描述如图 1 所示。

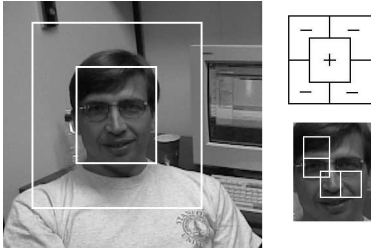


图 1 局部直方图描述

其中,目标区域和目标周围的背景区域用 2 个矩形框表示。中心的小框内是要跟踪的目标,定义为分类问题中的正样本。而目标 4 倍大小的外框包含了更多背景部分,它们被分割为和目标同样大小的 4 块,作为负样本。

此处提取的特征是直方图特征。这些直方图包含 2 种空间结构:水平和垂直方向上的局部叠加,即每个局部直方向量由其内部 2 个区域直方向量的叠加组成。这种叠加在某种程度上嵌入了一定的空间结构信息。而这些直方向图区域的位置可随机生成,也可按某些规则来决定。

为满足 Co-Training 的特征独立要求,必须从 2 种独立的特征空间来抽取直方向图特征。这里采用区域颜色直方向图和区域梯度方向直方向图两种不同的特征。其中,区域颜色直方向图主要由这个区域里所有像素的颜色分布所决定,而区域的梯度方向直方向图更多的取决于该区域的纹理信息,因此,这 2 种特征在较大程度上满足相互独立的要求。

梯度方向直方向图在计算时,首先计算该图像区域中每个点的梯度。假设某点为 (x, y) ,定义 x 轴方向的梯度为 $dx=I(x, y+1)-I(x, y-1)$, y 轴方向的梯度为 $dy=I(x+1, y)-I(x-1, y)$,则该点的梯度方向定义为 $\theta(x, y)=\arctan(dy/dx)$,再将该方向值的范围量化到 N 个区间(bin),统计整个区域每个点的梯度方向,即可得到该区域的梯度方向直方向图。

为加快特征提取的速度,本文采用积分直方向图技术^[6]。该技术利用类似积分图像的原理,在计算时,能有效减少重叠区域的重复计算,从而大大提高计算速度。

2.3 分类器设计

AdaBoost 作为集成学习的一种,由于其较好的泛化能力,在很多分类问题上取得了优异的效果。并且这种分类器的计算量较小,非常适合目标跟踪问题。AdaBoost 首先学习一组弱分类器,通过弱分类器的集成得到一个强性能的分类器。离散形式的 AdaBoost 的定义为

$$H(x)=\text{sign}(\sum_{t=1}^T \alpha_t h_t(x)) \quad (1)$$

其中, $h_t(x)$ 称为弱分类器,一般的形式为

$$h_t(x)=\begin{cases} 1 & \text{if } F_t(x) \geq T_t \\ -1 & \text{其他} \end{cases} \quad (2)$$

弱分类器的设计是 AdaBoost 方法的核心。对于传统的 Haar 形式的一维特征,只需要一个简单的域值比较即可定义一个简单的弱分类器。而对于向量形式的多维特征,比如此处的直方向图特征,则需要 SVM, KNN 等复杂一些的分类器。考虑到计算量,这里采用线性判别分析(linear discriminant analysis)作为弱分类器。

给定 2 类样本,在线性判别分析中可找到一个最优的投影向量 f 来划分这 2 类样本:

$$f = S_w^{-1}(u_+ - u_-) \quad (3)$$

其中, S_w 是 2 类样本的类内散度矩阵和; u_+ 和 u_- 是 2 类样本的均值。

$$S_w = \sum_{x_i \in D_+} (x_i - u_+)(x_i - u_+)^T + \sum_{x_j \in D_-} (x_j - u_-)(x_j - u_-)^T$$

在获得标记好的若干帧初始图像后,利用样本集可以训练出初始的 AdaBoost 分类器,训练过程如下:

输入 从初始 n 帧标记图像中抽取 n 个目标样本和 m 个背景样本 $(x_1, y_1), (x_2, y_2), \dots, (x_{m+n}, y_{m+n})$, x_i 表示一个样本,包含该样本区域提取出的所有直方向图特征集合,标识 $y_i=+1, -1$ 分别表示目标和背景

输出 将 T 次迭代之后的弱分类器进行加权得到最终的强分类器: $H(x)=\text{sign}(T(x))$, $T(x)=\sum_{t=1}^T \alpha_t h_t(x)$

初始化 每个样本对应的权值 $w_{t,i}=1/(m+n)$

for $t=1, 2, \dots, T$

(1)权值归一化 $w_{t,i} \leftarrow w_{t,i} / \sum_{j=1}^{m+n} w_{t,j}$

(2)对每个特征 j ,用式(3)训练出一个弱分类器 h_j ,分类器误差为 $\epsilon_j = \sum_{i=1}^{m+n} w_{t,i} / h_j(x_i) - y_i /$

(3)选择误差最小的分类器 h_t 为 h_t

(4)更新样本的权值 $w_{t+1,i} = w_{t,i} e^{\alpha_t h_t(x_i) - y_i}$, 其中, $\alpha_t = \frac{1}{2} \ln(\frac{1-\epsilon_t}{\epsilon_t})$

2.4 协同跟踪框架

对于新的图像帧,上一帧目标位置的周围是目标最可能重新出现的区域。因此,在上述区域内依次抽取备选目标样本,并加入到 2 个分类器中,计算对应的 $T(x)$ 值即可得到该备选目标的置信度。将区域内每个点的 $T(x)$ 值变换到 $[0,1]$ 区间内,可得 2 幅目标的置信分布图。

根据场景的一些特性,将 2 幅置信分布图以一定的权重进行加权。例如在光照变化较多的场景中,由于梯度特征的光照鲁棒性,可给予梯度置信分布图以较大的权重。在最终得到一个合成的目标置信分布图之后,挑选置信值最大的那个位置作为目标的当前位置。上述过程可通过遍历搜索或采用 mean-shift 之类的梯度算法来完成。

在实际跟踪问题中,为适应目标和背景的变化,必须不断地在线更新分类器。为避免自学习导致的分类器性能下降,本文采用 Co-Training 来协同更新 2 个 AdaBoost 分类器,即 2 个分类器将各自分类置信度最高和最低的几个样本作为正负样本交予对方做进一步更新。算法流程如下:

输入 视频帧 I_1, I_2, \dots

输出 每帧的目标位置 L_t

初始化 由标定好的前 n 帧图像训练出 2 个分类器 $H_1(x)$ 和 $H_2(x)$

在线跟踪: 从 $n+1$ 帧到视频结束

(1)在第 i 帧图像中,将上一帧目标位置 L_{i-1} 周围区域中的每个位置作为当前帧的目标备选位置

(2)对每个备选位置的目标样本,分别用 2 个分类器得出各自的置信值,进而得到 2 幅整个区域的目标位置置信分布图

(3)融合 2 个得出的置信分布图,得到目标的最终位置

(4)将 $H_1(x)$ 中置信度最高的几个样本作为正样本,置信度最低的几个样本作为负样本添加到 $H_2(x)$ 的训练集中,更新 $H_2(x)$ 分类器

(5)对 $H_2(x)$ 做类似操作,得到的样本用来更新 $H_1(x)$ 分类器

其中,(4)和(5)中的集成分类器 $H(x)$ 中权重 α_i 较小的 p 个弱分类器将被去除,同时新添加 p 个新的弱分类器,以此来达到分类器的更新。

3 实验与分析

实验采用常用的 Dudek 图像序列作为测试集。该序列由于包含运动目标的姿态、外表变化以及短暂的遮挡,场景的

光照变化等因素,已成为目标跟踪测试的一个常用测试集。该测试集的图像分辨率为 740 像素×480 像素。实验在 Intel Core2 3.5 GHz, 双核 CPU, 2 GB 内存的计算机上完成。

实验中,初始 10 帧目标位置由人工标定。提取直方图特征时,图像子区域的大小为 60×40,颜色直方图取 16 bin,梯度方向直方图取 8 bin。同时为了验证 Co-Training 框架的有效性,该算法与采用自学习框架的 Ensemble Tracker^[2]进行比较,实验结果如图 2 所示。

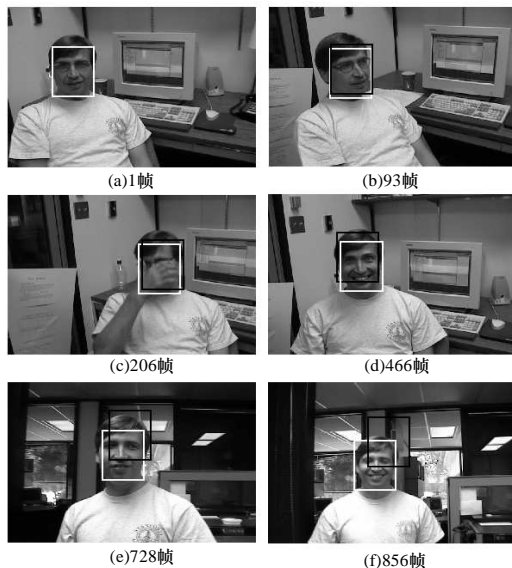


图 2 实验结果

其中,白框为本文提出方法,黑框为 Ensemble Tracker 结果。从实验结果可以看出,在测试序列的前几百帧中,无论本文提出的方法还是 Ensemble Tracker 基本上都能比较精确地定位目标。但随着跟踪过程的进行,目标和环境发生了一些变化,为捕捉这种变化,更多的目标样本被添加进入分类器进行重新训练。Ensemble Tracker 由于采用了自学习的策

(上接第 201 页)

为了获得较好的分类模型,在生成训练样例和测试样例的过程中进行了一定的限制,导致正确识别的指代个数减少。经过调整过滤条件,虽能适当提高召回率,但准确率和 F 值都有所下降,且召回率仍不够理想,还有提升的空间。

6 结束语

与其他系统相比,本系统具有如下特点:(1)系统性能好,得到了较高的 F 值;(2)实验平台和领域无关;(3)对不同种学习方法的兼容性较强,性能相对稳定。

在特征方面,虽然本系统选取了 12 个特征,但真正起到作用的仅有几个重要特征。本文采用 12 个常用的基本特征进行实验,取得了较好的结果。实验表明并不是所有的特征都能起到有用的作用,如距离特征,若按 Soon 等人的处理方法,用 SVM 甚至会降低系统性能,而仍存在召回率偏低的问题。所用的语料现仅只限于 MUC,缺乏与 ACE 实验平台上结果的对比。近 2 年的 ACL 的文献表明,语义特征在指代消解中有较好的应用。本文虽然也用到了语义特征,但取的是最基本的语义类别信息。

基于上述问题,下一步的研究方向为:(1)利用距离特征对训练样例的生成做一定的限制;(2)采用 ACE 的语料进行实验,同时获取有效的特征;(3)尝试从 WordNet 或 Web 中

略,从而导致了误差的累积,以至于到了第 728 帧之后已经出现了比较大的偏差,到第 856 帧时基本上已经丢失了目标。而本文提出的方法,由于结合了 Co-Training 框架,将样本更新带来的误差降到了最低,因此始终能比较稳定地跟踪到目标。在未经优化的 C++实现下,本文提出的方法能以 16 帧/s 运行,经过优化后能有望接近实时处理。

4 结束语

运动目标跟踪是很多计算机视觉应用的重要组成部分。实际跟踪过程中,为适应目标和场景的变化,跟踪方法需要不断更新模型。现有的跟踪方法通过自学习的框架来达到这个目的,由此带来误差的累积导致了跟踪算法的不稳定。本文采用 Co-Training 的框架,提出用 2 种独立的特征空间来描述目标和背景,在跟踪过程中,2 个特征空间各自对应的分类器协同工作,大幅提高跟踪的鲁棒性。

参考文献

- [1] Collins R T, Liu Yanxi. On-line Selection of Discriminative Tracking Features[C]//Proc. of the International Conference of Computer Vision '03. Nice, France: [s. n.], 2003.
- [2] Avidan S. Ensemble Tracking[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(2): 261-271.
- [3] Ross D, Lim Jongwoo, Yang Minghsuan. Adaptive Probabilistic Visual Tracking with Incremental Subspace Update[C]//Proc. of the 8th European Conference on Computer Vision. Prague, Czech Republic: [s. n.], 2004.
- [4] Chapelle O, Scholkopf B, Zien A. Semi-supervised Learning[M]. Cambridge, MA: MIT Press, 2006.
- [5] Blum A, Mitchell T. Combining Labeled and Unlabeled Data with Co-training[C]//Proc. of the 11th Annual Conference on Computational Learning Theory. Madison, USA: [s. n.], 1998.
- [6] Porikli F. Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces[C]//Proc. of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, CA, USA: [s. n.], 2005.

抽取更多的语义信息进行实验。

参考文献

- [1] 王厚峰. 指代消解的基本方法和实现技术[J]. 中文信息学报, 2002, (6): 9-17.
- [2] Wee Meng Soon, Hwee Tou N H T, Lim D C Y. A Machine Learning Approach to Coreference Resolution of Noun Phrase[J]. Computational Linguistics, 2001, 27(4): 521-544.
- [3] Vincent N, Cardie C. Improving Machine Learning Approaches to Coreference Resolution[C]//Proc. of the 40th Annual Meeting of the Association for Computational Linguistics. Philadelphia, PA, USA: [s. n.], 2002.
- [4] Taylor J S. 支持向量机导论[M]. 李国正, 王 猛, 曾华军, 译. 北京: 电子工业出版社, 2004.
- [5] Yang Xiaofeng, Su Jian, Zhou Guodong, et al. Improving Pronoun Resolution by Incorporating Coreferential Information of Candidates[C]//Proc. of ACL'04. Barcelona, Spain: [s. n.], 2004.
- [6] Zhou Guodong, Su Jian. A High-performance Coreference Resolution System Using a Constraint-based Multi-Agent Strategy[C]//Proc. of COLING'04. Geneva, Switzerland: [s. n.], 2004.

