

文章编号:1001-9081(2005)10-2351-03

快速人脸检测系统的设计与实现

吴瞰华,周昌乐

(厦门大学 计算机科学系,福建 厦门 361005)

(me_54@163.com)

摘要:为了快速而准确地在复杂背景的灰度或彩色图像中检测人脸,对实时人脸检测方法做了一些扩展并根据该方法设计实现了一个快速人脸检测系统,系统分为训练和检测两部分,训练方法为 AdaBoost 算法。测试表明该系统的性能较好。

关键词:人脸检测;矩形特征;AdaBoost;层叠分类器

中图分类号: TP391.4 **文献标识码:** A

Design and implementation of fast face detection system

WU Tun-hua, ZHOU Chang-le

(Department of Computer Science, Xiamen University, Xiamen Fujian 361005, China)

Abstract: In order to detect faces in gray or color images at high speed and accuracy, some extension to the technology of real-time face detection were made and a rapid face detection system was designed and realized. The training subsystem is based on AdaBoost algorithm. The experiment results show that the extension to Viola's technology is quite successful. The rapid face detection system is divided into two parts, training(or learning) subsystem and detecting subsystem.

Key words: face detection; rect-feature; AdaBoost; cascaded classifier

0 引言

运用 AdaBoost 算法^[1]训练分类器并利用这些分类器快速检测物体的方法的最大优点,是速度快而且准确(Viola 演示程序时在 PIII 700MHZ 的笔记本电脑上对 384×288 大小的动态图像进行检测,可以达到 15 帧/秒,而且检测率超过 90%)。

Viola 方法是一种有监督的统计模式识别方法。由于 Viola 方法还有许多亟待解决的问题,如:人脸旋转角度的适应性不够好;对图像灰度变化适应性不够好;检测率还不够理想等。另外,Viola 并未给出具体的实现细节。因此,作为一种探索、一种尝试,对 Viola 方法做了一点扩展并设计实现了一个快速人脸检测系统。实验表明该系统具有比较好的性能:对于 CMU 的比较公认的测试图像集,其中有 130 张图像,包含 507 张人脸,系统的检测率达到约 ,检测速度较以往的系统有较大的提高。

1 系统框架

系统框架如图 1 所示,分为训练和检测两大部分。

1) 训练部分:收集人脸样本和非人脸样本;从这两类样本集中提取所有矩形特征;矩形特征作为训练过程的输入变量,训练过程依据 AdaBoost 算法(自举算法的一个变种)展开,而训练目标是生成一个层叠分类器。层叠分类器由若干强分类器层层相连而成,每个强分类器由若干弱分类器构成,每个弱分类器对应一个被严格筛选得到的矩形特征。每个弱分类器和强分类器都有阈值的概念。

2) 检测部分:按某个策略从待检测图像中提取所有将被检测的子窗口;利用训练得到的层叠分类器对每个子窗口进行检测;对检测结果进行后处理,最后得到图像中可能包含的

所有人脸的坐标和范围。

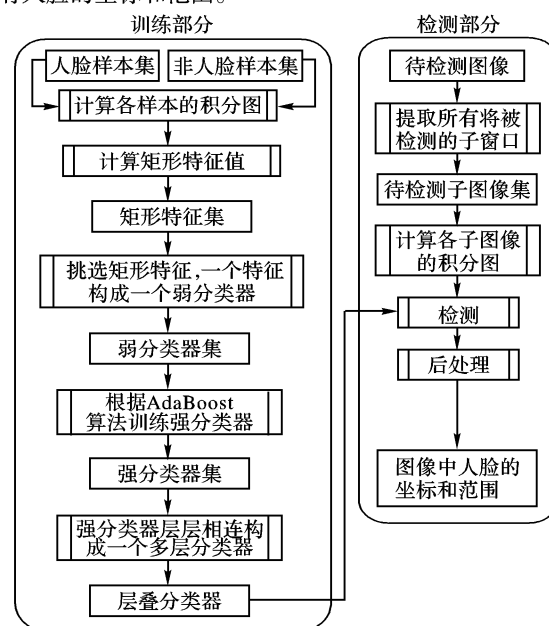


图1 系统框架图

2 矩形特征

矩形特征是一种简单特征,具有一定的人脸和非人脸区分性。训练前先从原始样本图像抽取矩形特征值,再由把矩形特征作为训练过程的输入变量,利用算法挑选出某些分类能力强的矩形特征分别构成各自对应的弱分类器,然后利用 AdaBoost 方法按目标驱动的方式将多个弱分类器组成一个强分类器。

如图 2 所示,左侧的人脸图像表示一张训练用的人脸样本

收稿日期:2005-04-18;修订日期:2005-06-30

作者简介:吴瞰华(1978-),男,福建泉州人,博士研究生,主要研究方向:人工智能技术及应用;周昌乐(1959-),男,江苏太仓人,教授,博士生导师,主要研究方向:计算机语言学、神经动力学和认知逻辑学等。

或者一张待检测图像;图上方两个包含黑白区域的矩形就代表矩形特征。矩形特征可以位于图像任意位置,大小也可以任意。图2展示这两个矩形特征在图像中的某两个位置。

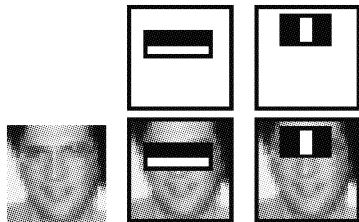


图2 矩形特征举例

2.1 矩形特征

矩形特征定义^[1,2]:

$$feature_i = \sum_{i \in \{1, \dots, N\}} \omega_i \cdot RecSum(r_i)$$

其中, $\{1, \dots, N\}$ 表示该特征由 N 个矩形构成, ω_i 为第 i 个矩形的权值, $RecSum(r_i)$ 为第 i 个矩形内所有像素值之和。

满足以上规定的矩形特征有无穷多个, Viola 对特征的构成进行了限制: 1) N 必须为 2; 2) 构成矩形特征的两个矩形对应的权值必须异号, 一正一负, 而且权值与矩形面积成反比。

在本系统中共采用了 14 个的矩形特征原型, 如图 3 所示。归纳起来, 它们分别属于以下三类抽象特征: 边缘特征、线特征、环绕特征。比起 Viola 给出的矩形特征原型, 这里已经加入旋转 45 度角的新矩形特征原型。从这些矩形特征原型可派生出丰富的矩形特征。

仅在 24×24 的窗口中就包含 117 941 个矩形特征, 如果每次计算特征值时都要分 2 次统计矩形内所有像素值之和, 那么计算量巨大。为此, 引入一种叫做“积分图”^[1,2] 的辅助工具, 每张待检测图像以及所有训练样本图像都可计算对应的积分图, 计算矩形特征值时只需 4 次查找积分图就可得解。

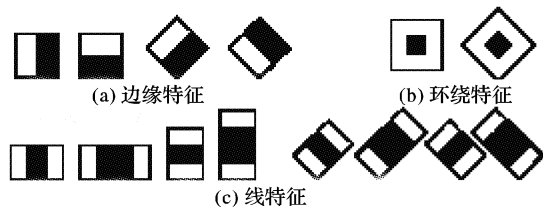


图3 矩形特征原型

3 训练分类器

矩形特征将作为训练过程的输入变量, 训练过程依据 AdaBoost 算法展开, 训练目标是生成一个层叠分类器。

3.1 层叠分类器的构成

为了能够显著地提高整个人脸检测过程的速度, 必须考虑从各个方面来降低计算成本。最直接的想法就是将一个复杂的分类器拆分为许多简化的分类器, 然后对这些简化了的分类器进行筛选, 组成一些较为复杂的分类器, 最后再把这些较为复杂的分类器层层相连^[1,2]。这些简化的分类器足够的小以至于可以在大约 20 条指令内就完成计算。这样, 系统的计算速度就可能迅速提高。

上述简化了的分类器就叫做弱分类器 (Weak Classifier)。在该系统中, 弱分类器与经过筛选得到的矩形特征一一对应, 弱分类器的原型是:

$$h_j(x) = \begin{cases} \alpha_1 & \text{当 } f_j(x) < \theta_j \\ \alpha_2 & \text{反之} \end{cases}$$

其中 x 为待检测窗口, $f_j(x)$ 为计算矩形特征值的函数, θ_j 为通过训练得到的特征值的阈值, 也称为弱分类器的阈值。

α_1, α_2 为两个小数, 值域为 $[-1, +1]$, 表示分类结果, 理想情况下分类完全正确, 那么 $|\alpha_1| = |\alpha_2| = 1$, 因为 $+1$ 表示人脸, -1 表示非人脸。

较为复杂的分类器就是强分类器, 由若干弱分类器组成。强分类器的原型是:

$$H_i(x) = \begin{cases} 1 & \text{当 } \left[\sum_{j=1}^n h_j(x) \right] > \varphi_i \\ 0 & \text{反之} \end{cases}$$

其中 x 为待检测窗口, $h_j(x)$ 为构成该强分类器的第 j 个弱分类器, φ_i 为该强分类器的阈值。 $H_i(x)$ 的判断结果为 1 和 0, 1 表示接受, 0 表示拒绝。

3.2 层叠分类器的检测率与误检率

假设有一个层叠分类器, 其各个强分类器的误检率为 f_1, f_2, \dots, f_n , 检测率为 d_1, d_2, \dots, d_n , 则整个层叠分类器的误检率

$$F = \prod_{i=1}^n f_i, \text{ 检测率 } D = \prod_{i=1}^n d_i.$$

3.3 训练方法

AdaBoost 方法是 Boost (自举) 方法的一个变种, 由 Freund 和 Schapire 于 1995 年提出^[9] (Freund 正是自举方法的提出者)。2001 年 Viola 扩展了该方法, 并把它用于物体检测 (不仅是人脸检测)^[1,2], 使它成为一种比较通用的物体检测方法。这个方法的基本思想是通过组合一组弱分类器来组成一个强大的分类器, 弱分类器由一个或多个非常简单的特征构成, 新的弱分类器的选择依赖于样本权值的改变。

层叠分类器的每一层都是用 AdaBoost 方法训练得到的。如果把基于 AdaBoost 的训练模块比作中央处理器的话, 那么所有样本的所有矩形特征值就是其输入, 而层叠分类器里的各个强分类器就是其输出。

下面对训练过程作了一个概括性的描述:

1) 确定系统目标误检率 F_{\max} , 各强分类器的最大误检率 f_{\max} , 最小检测率 d_{\min} , 并推断系统至少需要 $(n = \log_{f_{\max}} F_{\max})$ 个强分类器组成。

2) 确定训练系统需要多少人脸样本, 多少非人脸样本; 假设人脸样本需要 p 个, 非人脸样本需要 q 个。

3) 获取初始人脸样本集与非人脸样本集。

4) for $t = 1 : n$

训练一个强分类器 $H_t(x)$

组合前 t 个强分类器 H_1, H_2, \dots, H_t , 对人脸样本集进行验证, 淘汰被错误判断的人脸样本 (漏判), 并修改人脸样本的数量—— p 的值。

组合前 t 个强分类器 H_1, H_2, \dots, H_t , 对当前非人脸样本集进行验证, 淘汰被正确判断的非人脸样本, 并重新获取非人脸样本, 以补充非人脸样本集, 使其数量重新达到 q 个。在重新获取非人脸样本过程中, 也是组合前 t 个强分类器对候选非人脸样本进行验证, 只有被误判的候选非人脸样本才能被加入到非人脸样本集中。

5) 保存训练结果。

所谓验证就是利用前 t 个强分类器组成的临时层叠分类器对样本进行检测。在训练过程中, 需要不断更新非人脸样本集。在验证时有一部分非人脸样本会逐渐被筛选掉, 并不会通过所有层, 被筛选掉的非人脸样本对于下一层的训练已经没有价值了, 所以非人脸样本集是需要更新的。

4 检测人脸

4.1 传统检测方法的缺点

传统检测流程用一句话概括是: 按比例逐层缩小待检测

图像,形成“图像金字塔”;在“金字塔”中穷举待检测子窗口;把各个待检测子窗口作为分类器的输入,得到检测结果。传统方法的特点是检测窗口大小固定,图像不断被缩小。缺点在于如果待检测图像较大,缩小图像的变换需要占用大量时间,减慢检测速度。

4.2 本系统的检测方法

采用的方法是:待检测图像大小不变,逐层等比放大检测窗口,再用各层窗口遍历整张待检测图像。这样就提高了检测速度。

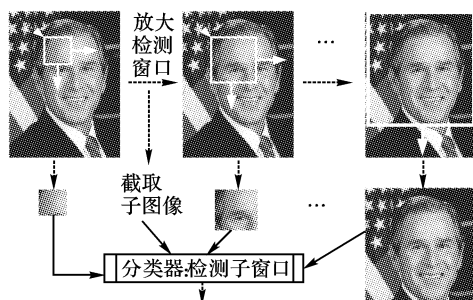


图4 本系统的检测流程

4.3 检测结果的后处理

检测结果可能存在结果窗口重叠问题。解决方法是:1)寻找可合并的结果窗口;2)合并重叠窗口(计算平均窗口)。

5 测试与分析

结合训练和检测两部分就构成起了一个完整的人脸检测系统。系统采用 MS VC7.0 结合 Intel OpenCV(机器视觉开发包)开发,运行环境是:CPU—AMD AthlonXP2200+、内存—512M DDR400、操作系统—WindowsXP Pro。

5.1 测试结果

1) 对 CMU 测试集(黑白图片)的检测结果:

表1 对 CMU 测试集的检测结果

检测出的人脸数	检测率	误检的子窗口数	误检率
478	$\frac{478}{507} \approx 94.3\%$	54	$\frac{54}{75081800} \approx 7.19 \times 10^{-7}$

2) 对彩色图片集的检测结果:

表2 对3张彩色图像的检测结果

序号	图像分辨率 V	检测率	误检的子窗口数	检测时间(秒)
1	384 × 288	100%	0	0.1
2	2204 × 1257	100%	2	9
3	1024 × 768	97.1%	0	2

5.2 比较与分析

5.2.1 性能比较

表3 各方法在 CMU 测试集上的性能比较

人脸检测方法	检测率	误检的子窗口数
扩展 Viola(本方法)	94.3%	54
Viola ^[1,2]	91.8%	167
神经网络 ^[3]	92.5%	862
相关信息量 ^[4]	98%	12758
Naive Bayes ^[5]	93%	88
SNoW ^[6]	94.8%	78
FLD ^[7]	93.6%	74

5.2.2 测试分析

1) 检测速度:在分辨率为 384 × 288 的图像上,本系统的检测时间为 1/10 秒,比起 Paul Viola 给出的 1/15 秒来说已经很接近,也就是说本系统的检测速度离 Viola 所谓的“实时检测”已经很接近了。再比较 Rowley 提出的神经网络的方法,在同样的分辨率下,其检测时间为 1 分钟 ~ 3 分钟,速度很慢。

2) 检测率和误检率:对于 CMU 测试集来说,94.3% 的检测率达到了主流方法的水平,而误检率比以往其他方法要低。CMU 测试集的图像质量较高,因此检测率较高。

3) 鲁棒性:检测率跟图像质量、人脸是否旋转有很大关系;误检率比较不受这两个因素的影响。测试表明系统对人脸的上下俯仰、左右侧转适应性较高,适应度达到了左右;系统对人脸的左右摆动适应性很差,适应度只有左右。为了解决这个问题,必须继续深入研究,提出根本的解决方法。不过,本系统可以检测任意格式的彩色或灰度图像,而基于肤色模型的方法对灰度图像无能为力。总的来看,鲁棒性可以接受,但须要提高。

6 结语

本文扩展了 Viola 方法,并根据扩展后的方法设计、实现一个人脸检测系统。对 Viola 方法的扩展主要包括:1)引入 45° 倾角的矩形特征,扩大了训练的范围,提高了系统的检测率,降低了误检率。2)采用与众不同的检测策略,避免直接对图像进行缩放变换,减小了计算量,提高了检测速度。

系统还存在一些不足,主要是鲁棒性还不够理想,尤其是对旋转的人脸适应性较差。检测率和检测速度也还必须继续提高。当前,已有将 Viola 方法应用于车牌检测、人体检测等一般物体检测中,也取得了不错的成效,因此将 Viola 方法应用于更广的领域,也是今后研究的内容之一。

参考文献:

- [1] VIOLA P, JONES M. Robust Real-time Object Detection[J]. International Journal of Computer Vision, 2004, 57(2): 137-154.
- [2] VIOLA P, JONES M. Rapid Object Detection using a Boosted Cascade of Simple Features[A]. Proceedings of the IEEE CVPR[C], USA: IEEE, 2001. 511-518.
- [3] ROWLEY HA, SHUMET B, KANADE T. Neural Network-Based Face Detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(1): 23-28.
- [4] COLMENAREZ AJ, HUANG TS. Face Detection with Information-based maximum discrimination[A]. Proceedings of the IEEE CVPR[C], USA: IEEE, 1997. 782-787.
- [5] SCHNEIDERMAN H, KANADE T. Probabilistic Modeling of Local Appearance and Spatial Relationships for Objects Recognition[A]. Proceedings of the IEEE CVPR[C], USA: IEEE, 1998. 45-51.
- [6] ROTH D, YANG MH, AHUJAA N. SNoW - based Face Detector[J]. Neural Information Processing Systems(NIPS), 1999, (12): 34-40.
- [7] YANG MH, AHUJAA N, KRIEGMAN D. Mixtures of Linear Subspaces for Face Detection[A]. Proceedings of the IEEE ICCV[C], USA: IEEE, 1999. 67-72.
- [8] YANG MH, AHUJAA N, KRIEGMAN D. Detecting Faces in Images: A Survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(1): 54-61.
- [9] SCHAPIRE FRE. A Short Introduction to Boosting[J]. Journal of Japanese Society for Artificial Intelligence, 1999, 14(5): 771-780.
- [10] 梁路宏. 人脸检测与跟踪[D]. 北京: 清华大学计算机科学系, 2002.