

基于改进 AdaBoost 算法的人耳检测与跟踪

张 惟 穆志纯 袁 立

(北京科技大学信息工程学院, 北京 100083)

摘要 人耳检测是人耳识别系统的第一个环节。在比较已有的人耳检测方法的基础上,介绍了一种复杂背景下的快速人耳检测与跟踪的方法。该方法主要分为两个阶段,离线级联分类器训练阶段和在线检测阶段。在离线训练阶段,首先结合人耳轮廓清晰,凹凸有致的特点,采用扩充后的 haar-like 型特征,依最近邻法则构造出弱分类器空间,然后根据经验选择 GAB 算法训练出强分类器,最后将多个强分类器级联成多层人耳检测器。在线检测阶段,为提高检测率,本文采用了调整分类器阈值和缩放检测子窗口的策略。最终检测器在 CAS-PEAL 人脸库上测试,检测率达到 98% 以上;在 PIV 1.7GHz 的 PC 上对普通 CMOS 摄像头输入的 320×240dpi 视频进行人耳跟踪,速度可达 6~7fps。实验结果表明,本文的人耳检测方法具有较好的实时性和一定的鲁棒性。

关键词 人耳检测 Gentle AdaBoost haar-like 型特征 级联分类器

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1006-8961(2007)02-0222-06

Fast Ear Detection and Tracking Based on Improved AdaBoost Algorithm

ZHANG Wei, MU Zhi-chun, YUAN Li

(Information Engineering School, University of Science & Technology Beijing, Beijing 100083)

Abstract Ear detection is the first step of an ear recognition system. On the basis of comparing to other existing approaches, this paper introduces a kind of fast ear detecting and tracking approach under the complex background, which has two stages: off-line cascaded classifier training and on-line ear detection. In the stage of off-line training, considering the vivid contour, the concave and convex of the ear, we apply the extended haar-like features to construct the space of the weak classifiers using the nearest neighbor norms. And then we choose the GAB algorithm with experience to train the strong classifiers which form the cascaded multi-layer ear detector. In the stage of the on-line detecting, we adopt the methods of adjusting the threshold of the strong classifiers and zooming in out the detecting sub-windows for speeding up while keeping the size of the original image. The testing experiments on the face database of CAS-PEAL result an upwards of 98% hit rate, and on the 320×240 dpi video inputted by the CMOS camera using PIV 1.7GHz PC result a speed of 6~7fps, which shows that the proposed method is significantly efficient and robust.

Keywords ear detection, Gentle AdaBoost, haar-like features, cascaded classifier

1 引言

本文的研究背景基于人耳识别系统。人耳检测作为人耳识别的第一个环节,其效果直接影响到后续工作。依据目前人耳检测研究的现状,若将人耳视为 2 维空间的平面图像,由于受光照、姿态等因素的影响,难度可想而知;若将人耳视为刚体,其在 3

维空间中的转动又会产生形状各异的平面投影;在现实世界中,人耳的颜色、形状更因种族、地域、年龄的各异而千差万别,此外光照、饰物、遮挡等又为检测带来种种不确定因素。

目前已有的人耳检测方法主要有以下几种:第 1 种是基于肤色统计模型的动态人耳检测方法^[1],该方法首先根据自适应背景模型从彩色序列图像中提取出运动物体,再由几何特征判断是否为人,若是

基金项目:国家自然科学基金项目(60573058,60375002);北京市教委重点学科共建项目(XK100080431)

收稿日期:2006-10-16;**改回日期:**2006-11-06

第一作者简介:张惟(1981~),女,北京科技大学控制理论与控制工程专业在读硕士研究生。研究方向为机器视觉、模式识别、目标跟踪。E-mail:mydream_cv@yahoo.com.cn

则运用色彩信息从中分割出肤色区域,然后用训练得到的人耳模板与该肤色区域内的人耳进行匹配,实现了人耳的动态跟踪;第2种是融合肤色及灰度轮廓信息检测跟踪序列图像中人耳的方法^[2],该方法首先在肤色概率分布的基础上,通过获取前几帧图像中肤色区域的信息,应用改进的 CAMSHIFT 算法对包含人耳的侧面人脸进行粗定位,进而根据耳朵自身丰富的轮廓信息,在假设人耳位于侧面人脸左上角位置的情况下,应用边缘检测和轮廓拟合进一步精确定位人耳。第3种是利用图像形状指数的平均直方图离线建立人耳模板^[3],然后在线对图像进行边缘检测、膨胀、模版匹配等处理,最终实现静态侧面人脸图像中的人耳检测。这些方法充分利用了人耳与侧面人脸之间颜色和位置的关系,同时结合人耳椭圆形的外耳轮廓和凹凸有致的内耳沟回信息,实现了正常光照和较复杂背景下的单人耳检测与跟踪。

但是这些方法均有其局限性。第1种方法只能应用于运动人耳的检测,且图像中必须包含整个人体;第2种方法只能检测动态图像,且对光照十分敏感,侧面人脸中人耳的位置对检测效果也有很大影响;第3种方法只能处理背景简单的静态图片中单人耳检测问题。

综上,借助侧面人脸定位人耳虽然简单易行,但是肤色信息对光照十分敏感,人耳的形状和相对于侧面人脸的位置又因人而异,如果人脸信息丢失,人耳检测也便失效;且这些方法只适用于实验室状态下的人耳检测,并未考虑人耳的姿态、饰物、遮挡等因素的影响,一旦将其置于真实世界复杂多变的情况下检测人耳,多样的背景、丰富的种族、地域特征,姿态、年龄、性别、数量等都将对系统的性能造成极大影响。如何运用人耳自身的特点,将其独立于其他生物特征进行检测,并使其对真实复杂的环境具有一定的适应性成为人耳检测研究的关键。考虑到人耳自身轮廓鲜明,线条丰富,形态各异,背景又千差万别,人为地寻找不同人耳之间,人耳与非人耳之间的差异和共性十分困难,而统计学习的方法能够借助机器对大量样本进行统计计算,便于从中找出规律,挑选出合适的特征进行分类,因此选用了统计学习方法中已成功应用于人脸检测的 AdaBoost 算法。

AdaBoost 算法最早源于 Schapire 在 1989 年提出的 Boosting(自举)算法^[4],它是一种能够“Boost”任意给定学习算法精度的普适方法。1995 年 Freund 和 Schapire 又对其进行了改进,形成了最初的 AdaBoost(Adaptive Boosting)算法^[5],之后又出现

了很多基于此方法的变形,主要有 GAB(Gentle AdaBoost)、DAB(Discrete AdaBoost)、RAB(Real AdaBoost)和 LAB(Logit Boost)。AdaBoost 算法的核心是通过调整样本分布和弱分类器权值,自动地从弱分类器空间中筛选出若干关键的弱分类器,并通过某种方式整合为一个强分类器。2001 年 Viola 等人^[6]又在 AdaBoost 算法的基础上,通过引入 haar-like 型特征和积分图像法(Integral Image),设计了级联式(Cascaded)人脸检测器,在人脸检测系统的检测率和检测速度上均迈出了关键的一步。之后 Rainer 等人^[7]又对系统使用的 haar-like 型特征进行了扩展,使该人脸检测系统具有更好的检测效果。

本文的贡献在于将机器学习引入到人耳检测系统中,通过使用扩展的 haar-like 特征来表达人耳的点线特征,并参照已有的实验数据对比结果选择 Gentle AdaBoost 算法训练出一组强分类器,以级联的方式将他们整合起来,构造出多层人耳检测器。为了加快检测速度,一方面实验通过调整分类器的阈值,令分类器前两层即排除大量疑似非人耳子窗口,仅留下疑似人耳子窗口参与后续分类,另一方面采用固定被检图像大小,缩放检测子窗口的方法对图像遍历,得到检测目标。

2 haar-like 型特征及积分图像法

2.1 haar-like 型特征

haar-like 型特征是 Viola 等人^[6]在其人脸检测系统中引入的一种简单矩形特征,因类似于 haar 小波而得名,如图 1 所示。haar-like 型特征的定义是在图像子窗口中黑色矩形和白色矩形对应区域内像素灰度值总和的差。从图中可以看出,这种矩形特征能反映检测对象局部特征的灰度变化,比如耳窝及其周围等。

由于人耳本身外耳轮廓清晰成椭圆形,内耳多沟回,且线条丰富,凹凸有致,若用椭圆来做轮廓拟合和

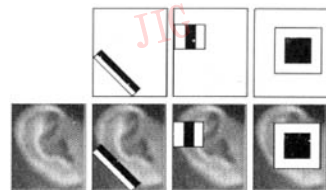


图1 图像子窗口中的矩形特征

Fig.1 Rectangle features in the sub-window of image

特征提取十分合适,但椭圆形状变化丰富,计算复杂,一旦特征数量巨大,计算量将不可想象。而用矩形来逼近局部特征不仅能够很好地拟合边缘、线条、点等特征,而且计算简单,极易推广。又由于耳朵形状线条多倾斜,因此实验参照 Rainer 等人^[7]对 Viola 方法的改进,引入了 45° 旋转矩形特征,并在此基础上选取了扩展后的 14 种矩形特征原形,分别表示边缘、线形和中心特征,如图 2 所示。通过改变这些矩形特征原形的边长还可派生出更加丰富的矩形特征。

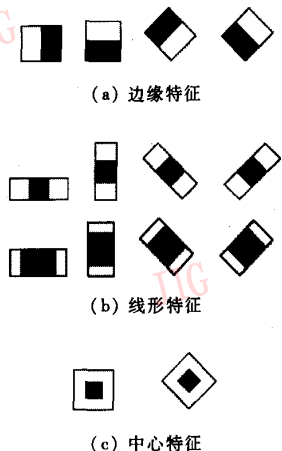


图 2 haar-like 型特征原形

Fig. 2 Feature prototypes of simple haar-like

2.2 积分图像法

一幅图像中的矩形特征数量是巨大的,以边缘特征中第 1 个矩形特征为例,在一幅 25 × 30 大小的窗口中由此原形派生出的特征数就为 73 125 个,可见一幅图像中的特征数要远远大于其像素数。因此本文采用了 Viola 等人^[6]提出的积分图像的方法来快速计算矩形特征。

采用这种方法在计算矩形特征时不必每次重新统计矩形内像素灰度值的和,而只需索引矩形几个相应点的积分图像值即可计算出矩形特征值,计算时间也不会随矩形大小的变化而变化。

3 多层分类器的构成

3.1 AdaBoost 算法

AdaBoost 算法之所以被称为自适应的 (Adaptive) Boosting 算法,是因为其能够对弱分类器集合中分类性能好的弱分类器给予较高权重,而对分

类性能差的弱分类器给予较低权重,进而在巨大的弱分类器空间中挑选出若干关键分类器,整合为一个强分类器,整合的规则是由若干弱分类器的性能决定的。算法的每一轮迭代过程中,样本的权值都会受错分样本权值之和的影响,若样本被分对,在下一轮迭代时该样本权值便降低,若样本被分错,在下一轮迭代时该样本权值就升高。

本文采用简单的最近邻法则作为弱学习方法,得到弱分类器,因此不必事先构造弱分类器集合,而是在特征筛选的过程中产生弱分类器,因此每一个弱分类器都与筛选得到的矩形特征一一对应。

根据 Rainer 等人对变形的几种 AdaBoost 算法的测试和比较,GAB (Gentle AdaBoost) 算法在分类性能上不仅是最优的,也是最快的,因此本文选用这种方法来训练强分类器,算法如下^[7]:

- (1) 给定 N 个样本 $(x_1, y_1), \dots, (x_N, y_N)$, 其中 $x \in \mathcal{X}, y_i \in \{-1, 1\}$ 。
- (2) 初始化权值 $w_i = 1/N, i = 1, \dots, N$ 。
- (3) 迭代 for $m = 1, \dots, M$:
 - (a) 通过 w_i 设置 y_i 到 x_i 的加权最小平方,调整退化函数 $f_m(x)$;
 - (b) 计算 $err_m = E_w [l_{(y \neq f_m(x))}]$, $c_m = \log((1 - err_m)/err_m)$;
 - (c) 令 $w_i \leftarrow w_i \exp(-y_i f_m(x_i)), i = 1, \dots, N$, 重新归一化权值使 $\sum_i w_i = 1$ 。

- (4) 输出强分类器: $\text{sign} \left[\sum_{m=1}^M f_m(x) \right]$ 。

3.2 级联式多层分类器

采用 Viola 等人在人脸检测系统中采用的方法,将 AdaBoost 算法训练得到的多个强分类器级联^[6],构成多层分类器,级联方式如图 3 所示。

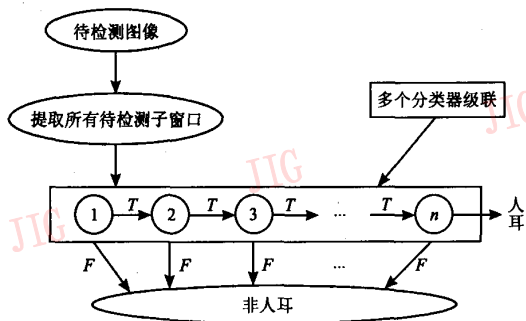


图 3 级联式多层分类器结构

Fig. 3 Cascaded classifier structure

图中 T 为检测率, F 为误检率, 其中每一层都对应一个 AdaBoost 算法训练得到的强分类器。越靠后的层弱分类器越多, 分类性能也越强。检测时图像子窗口只有通过所有层才能被判断为是正例, 若在其中任一层被判断为反例即被抛弃。

4 实 验

4.1 人耳样本集

(1) 人耳正样本

训练用正样本主要来自本实验室拍摄的人耳库, 从图像库 2 和彩色图像库中选取部分图像割取人耳, 另一部分取自中科院 CAS-PEAL 人脸库的 POSE 部分, 从中挑选出合适的侧面人脸图片割取人耳。如图 4 所示。由于样本数量不足, 对割取的左右耳样本均做了镜像处理, 得到左右耳正样本各约 5 000 幅, 共 10 000 个正样本。



图 4 训练用正样本
Fig. 4 Positive samples for training

参照人脸在空间中的姿态^[8], 若以摄像机垂直于人脸侧面为 0° , 可以把人耳的姿态分为 3 种, 即绕垂直轴旋转(摇摆)、绕视平面法向轴旋转(倾斜)和绕水平轴旋转(俯仰), 如图 5 所示。由于样本库的限制, 这里只考虑了左右耳部分姿态的样本, 姿态

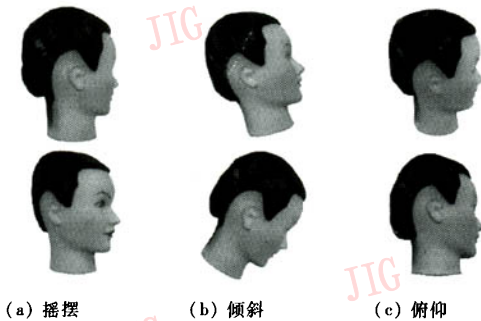


图 5 侧面人脸在空间中的旋转
Fig. 5 Rotation of the profile in the space

范围包括摇摆 $[-45^\circ, +45^\circ]$, 倾斜 $[-20^\circ, +20^\circ]$, 俯仰 $[-0^\circ, +0^\circ]$, 其中正/负号代表左/右或上/下, 有光照变化, 饰物和轻微遮挡(遮挡面积不超过人耳面积的 30%)。

考虑到人耳识别的最佳分辨率为 20×32 dpi, 并参考 Viola 人脸检测实验中正样本的大小 24×24 dpi, 同时尽量使样本包含完整的人耳, 这里选择以 25×30 及其倍数大小的矩形框从图像中分割出人耳, 最后统一成 25×30 大小的图片。另外, 为帮助提高检测效果, 样本图像中保留了少量毛发、皮肤等信息。

(2) 人耳负样本

训练用负样本主要来自自己拍摄的校园、实验室场景, 网络上下载的风景、动植物、建筑等的图片, 另一部分取自中科院 CAS-PEAL 人脸库中有光照变化和饰物的正面人脸, 共约 10 000 幅, 图像尺寸均大于 50×60 dpi。

(3) 人耳测试样本

由于缺少公开的较为标准的人耳测试库, 本文的测试样本一部分来自中科院 CAS-PEAL 人脸库中 POSE 部分的侧面人脸(未用来做正负样本), 另一部分取自 CMU 的 profile_images 侧面人脸库中的图片, 测试时去除了库中人耳偏转角度超出训练样本范围较大、有严重遮挡, 以及耳朵太小的图片, 同时还选取了一部分自己拍摄的照片及网上下载的图片。

4.2 人耳检测过程

(1) 检测预处理

① 对输入图像做滤波和光照补偿, 这里采用较为常用的中值滤波和直方图均衡化的方法。

② 检测窗口的初始值大小与训练用的正样本大小一致, 并设置检测窗口放大比例为 1.2。

(2) 检测方法及其加速策略

采用 Viola 人脸检测系统中使用的快速检测方法, 令被检图像大小不变, 逐层等比放大检测子窗口, 并用各层子窗口遍历被检图像。这样在每次缩放检测子窗口时只需重新计算检测窗口的长宽即可, 大大降低了计算量。

另外为加快检测速率, 本文对各层分类器的阈值进行了设置, 使多层分类器对图像进行检测时, 先令前两层强分类器快速排除大量疑似非人耳子窗口, 再用后面的强分类器对剩余的疑似人耳子窗口做进一步检测。

4.3 实验结果

实验得到的人耳检测器共 18 层,467 个特征,在一台 PIV 2.7GHz、512RAM 的 PC 上共训练约 16 天。检测过程是在 PIV 1.7GHz、256RAM 的 PC 上进行的。测试结果如下:

(1) 从中科院 CAS-PEAL 人脸库中 POSE 部分随机取出 166 幅人耳姿态为摇摆 $\pm 45^\circ$, 倾斜 $[-20^\circ, +20^\circ]$, 俯仰 $[-0^\circ, +0^\circ]$ 的侧面人脸图片,检测率在 98% 以上,误报 26 个。如图 6 所示。



图 6 CAS-PEAL 人脸库检测结果

Fig. 6 Detecting results on the face database of CAS-PEAL

(2) 从 CMU 的 profile_images 侧面人脸库中取出人耳偏转角度在训练样本范围内或稍大、有轻微遮挡,以及耳朵尺寸大于或略小于 $25 \times 30\text{dpi}$ 的图片共 70 张,有 26 张检测到耳,误报 6 个,测试效果不理想。其原因主要有以下几点:① 该图像库不针对人耳检测,因此图像中的人耳大多有严重遮挡,比如头发、饰物等,且偏转角度和姿态范围均超出了本文实验训练样本覆盖的范围;② 图像库中的人耳大多取自亚洲以外其他国家人的耳朵,除了形态上有差别外,其头发和肤色均与亚洲人有很大区别,且人物的年龄各异,装饰不同,而本文实验用的训练样本只包含了中国人的耳朵图片。测试的结果如图 7 所示。



图 7 CPU-PIE 侧面人脸库中的人耳检测结果

Fig. 7 Detecting results on the CMU-PIE database

(3) 实验从网上下载了若干图片,同时也自己动手拍摄了一些照片,作为检测样本,其中包含了多人耳检测的情况,检测结果如图 8 所示。

(4) 使用普通的 CMOS 摄像头,分辨率为 $320 \times 240\text{dpi}$,检测速度 $6 \sim 7\text{fps}$,能够满足实时检测的要求。如图 9 所示。

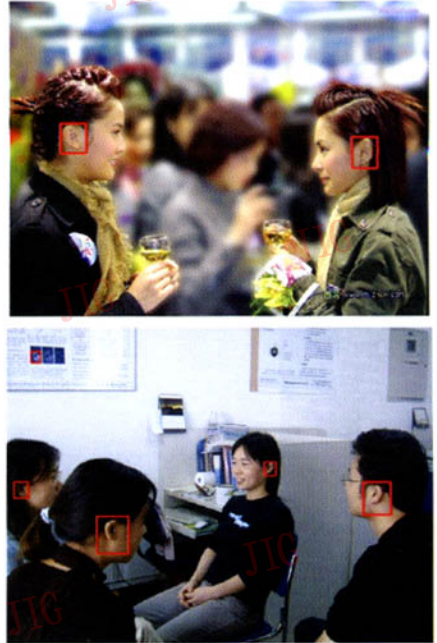


图 8 网上下载和自己拍摄的照片中人耳检测结果

Fig.8 Detecting results on the web resources and the photos of our taking



图 9 摄像头跟踪人耳的情况

Fig.9 Detecting results using CMOS camera

5 结论

本文构造了一个基于 AdaBoost 算法的级联式人耳检测系统,不仅克服了已有人耳检测方法对肤

色和人耳人脸相对位置的依赖,而且能够处理较复杂环境中的多人耳检测问题,对各种姿态、形状的左右人耳均较以往有较好的适应性;对人耳的动态实时检测在跟踪精度上好于已有方法,快速性上基本一致。实验选择了扩充后的 haar-like 型特征集,并在此基础上运用 GAB 方法构造出级联式的人耳检测器;检测阶段为弥补图像传输过程中带来的干扰,加入了预处理,同时为更好地实现实时性,采用了一些加速策略。从实验结果上看,AdaBoost 方法对样本的依赖性很强,因此人耳图像库的不完善,包括样本数量有限,且缺少各种姿态以及不同年龄、种族人的耳朵样本等,使得实验构造的系统存在一定的局限性;另一方面训练得到的分类器的效果会随着特征、样本和弱分类器方法选取的不同而产生很大差异,如何针对人耳不对称的特点构造出新的特征,选择合适的分类方法,以及如何训练左右耳样本,从而得到更有效的人耳检测器,都需要在将来做进一步的尝试和完善。整个系统在 Windows XP 系统 Visual C++ 6.0 环境下实现。

参考文献 (References)

- 1 Wang Ming, Mu Zhi-chun. Dynamic ear detection based on skin color model[J]. Control and Automation, 2005, 21(8-1):51~53. [王敏,穆志纯. 基于肤色模型的人耳检测系统[J]. 微计算机信息, 2005, 21(8-1):51~53.]
- 2 Zhao Y. Real-time Ear Tracking Based on Both Skin-color and Contour Information [D]. Beijing: University of Science and Technology, 2005. [赵祎. 基于肤色及轮廓信息的人耳实时检测与跟踪[D]. 北京:北京科技大学, 2005.]
- 3 Hui Chen, Bir Bhanu. Human ear detection from side face range images[A]. In: Proceedings of the 17th ICPR [C], Cambridge, UK, 2004, 3: 574~577.
- 4 Schapire R E. The strength of weak learnability [J]. Machine Learning, 1990, 5(2): 197~227.
- 5 Freund Y, Schapire R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Journal of Computer and System Sciences, 1997, 55(1): 119~139.
- 6 Paul Viola, Michael Jones. Robust Real-time Object Detection[R]. TR CRL 2001/01, Cambridge, UK: Cambridge Research Laboratory, 2001.
- 7 Lienhart Rainer, Kuranov Alexander, Pisarevsky vladim. Empirical analysis of detection cascades of boosted classifiers for rapid object detection [A]. In: Proceedings of the 25th German Pattern Recognition Symposium [C], Magdeburg, 2003: 297~304.
- 8 Wu Bo, Huang Chang, Ai Hai-zhuo. A multi-view face detection based on real adaBoost algorithm[J]. Journal of Computer Research and Development, 2005, 42(9): 1612~1621. [武勃,黄畅,艾海舟. 基于连续 Adaboost 算法的多视角人脸检测[J]. 计算机研究与发展, 2005, 42(9):1612~1621.]