

# 基于 SVM 的三阶段人脸检测方法的研究与应用

袁芬萍, 季桂树

(中南大学 信息科学与工程学院, 湖南 长沙 410083)

**摘 要:** 为了提高人脸的检测速度, 描述了一种基于 SVM 的三阶段人脸检测方法, 加速了复杂背景下人脸的检测。在该方法中, 通过前两个阶段的瀑布型分类器将大量的简单非人脸模块快速过滤掉, 并在最后的第三个阶段采用了非线性 SVM 分类器来更高效地区分  $24 \times 24$  像素的人脸或者非人脸图像。

**关键词:** 人脸检测; 特征检测; 瀑布型分类器; SVM

**中图分类号:** TP391.41

**文献标识码:** A

**文章编号:** 1673-629X(2007)09-0133-03

## Research and Application about Three-Stage Face Detection Method Based on SVM

YUAN Fen-ping, JI Gui-shu

(College of Info. Sci. & Eng., Central South Univ., Changsha 410083, China)

**Abstract:** In this paper, present a three-stage method to speed up an SVM-based face detection system. In this proposed system, a large number of simple non-face patterns are rejected quickly by two first stage cascaded classifiers using flexible sizes of analyzing Windows while the last stage uses a non linear SVM classifier to robustly classify complex  $24 \times 24$  pixel patterns as either faces or non-faces.

**Key words:** face detection; feature tracking; cascade classifier; SVM

### 0 引言

人脸检测(Face Detection)是指在输入图像中确定所有人脸(如果存在)的位置、大小、位姿的过程。人脸检测作为人脸信息处理中的一项关键技术, 近年来成为模式识别与计算机视觉领域内一项受到普遍重视、研究十分活跃的课题。

人脸检测问题最初来源于人脸识别(Face Recognition)。人脸识别的研究可以追溯到 20 世纪六、七十年代, 经过几十年的曲折发展已日趋成熟<sup>[1,2]</sup>。人脸检测是自动人脸识别系统中的一个关键环节, 但是早期的人脸识别研究<sup>[3]</sup>主要针对具有较强约束条件的人脸图像(如无背景的图像), 往往假设人脸位置已知或很容易获得, 因此人脸检测问题并未受到重视。近几年随着电子商务等应用的发展, 人脸识别成为最有潜力的生物身份验证手段, 这种应用背景要求自动人脸识别系统能够对一般环境图像具有一定的适应能力, 由此所面临的一系列问题使得人脸检测开始作为一个

独立的课题受到研究者的重视。今天, 人脸检测的应用背景已经远远超出了人脸识别系统的范畴, 在基于内容的检索、数字视频处理、视觉监测等方面有着重要的应用价值。

人脸检测研究具有重要的学术价值。人脸是一类具有相当复杂的细节变化的自然结构目标, 此类目标的检测问题的挑战性在于<sup>[4]</sup>:

(1) 人脸由于外貌、表情、肤色等不同, 具有模式的可变性;

(2) 一般意义下的人脸上, 可能存在眼镜、胡须等附属物;

(3) 作为三维物体的人脸的影像不可避免地受由光照产生的阴影的影响。因此, 如果能够找到解决这些问题的方法, 成功构造出人脸检测与跟踪系统, 将为解决其它类似的复杂模式检测问题提供重要的启示。

### 1 三层人脸检测系统

三层人脸检测系统包含三个阶段, 并被  $24 \times 24$  像素的窗口分为人脸区域和非人脸区域。为了检测出不同尺寸和不同位置上的人脸, 用 1.25 的影响因子作用在待检测的图像的每个区域和位置。

该检测系统的缩略图如图 1(三层人脸检测系统)

收稿日期: 2006-11-16

作者简介: 袁芬萍(1980-), 女, 陕西渭南人, 硕士研究生, 研究方向为高级数字图像处理; 季桂树, 副教授, 研究方向为高级数字图像处理。

所示。

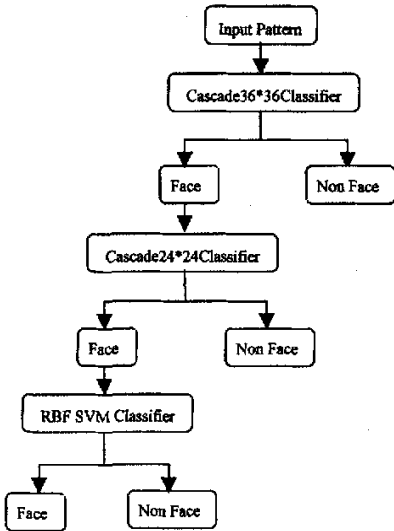


图 1 三层人脸检测系统

第一阶段采用了一个灵活可变的分类器,使用该分类器在  $36 \times 36$  像素大小的输入窗口中评估出人脸的候选区域。希望通过该分类器原样不变地使人脸区域显示在一个  $24 \times 24$  像素的窗口中。也就是说,如果一个  $36 \times 36$  像素的窗口检测到了某些人脸的存在,那么将选择出  $12 \times 12 = 144$  像素大小的最可能的人脸区域,并进入到该检测系统的下一个阶段。

第二阶段采用的分类器是一个  $24 \times 24$  像素的窗口,使用该分类器从第一阶段选定的区域中进一步探测出人脸的候选区域,并在进入最后一个阶段的分类器之前,尽可能多地过滤掉非人脸区域。

前两个阶段瀑布型的分类器类似于文献[5]中所描述的系统,但有所不同的是,三层人脸检测系统是由 Haar wavelet 特征集中的一个最佳子集构成,该子集在 AdaBoost 算法中被用来创建每一层的分类器,并在瀑布型结构的分类器中结合成层分类器。通过运用瀑布结构,分类器的复杂度在很大程度上就可以适应输入模式的复杂度了。在一、二阶段区域的选择和瀑布型分类器的快速评估下,大大提高了滤除非人脸区域的速度。

最后一个阶段的分类器是一个非线性 SVM 分类器<sup>[6]</sup>。在第二阶段的分类器中通过 AdaBoost 学习机选择出的特征被重新用作典型模板。确切地说,对于每一个  $24 \times 24$  像素的输入窗口,求出这些特征值,并使用 0 和 1 规范化表示它们,为基于 SVM 分类器的 RBF 形成一个特征向量。在实验中,因为仅仅使用了 200 个特征,所以,比起利用基于像素的模板要快得多。

## 2 人脸特征的选择

人脸特征是进行人脸检测的重要依据。和单个的像素比较,用特征进行检测有很多原因,其中的一个原因是特征本身就包含了部分局部信息,而这些局部信息知识是很难通过有限的训练数据得到的。采用人脸特征而不是单个像素的另外一个好处就是基于特征的检测<sup>[7]</sup>要比基于像素的检测快得多。因此,算法采用了一组类似于哈尔小波变换的滤波器来计算图像的特征值,如图 2 所示。图中包括 3 种特征:双矩形特征 (two-rectangle feature),三矩形特征 (three-rectangle feature) 和四矩形特征 (four-rectangle feature)。分别对应图中的 A, B, C 和 D。其中双矩形特征定义为图中相邻两个矩形内像素颜色之和的差。三矩形特征定义为两边的两个矩形减去中间的矩形中像素的颜色值之和。四矩形特征定义为对角线上矩形内像素颜色之和的差。

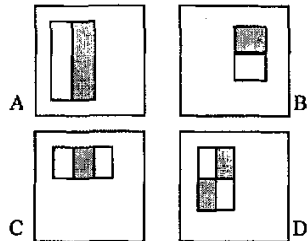


图 2 人脸矩形特征的定义

为了快速地在不同尺度下计算图像的特征,算法引入了积分图 (integral image) 的概念,如图 3 所示。积分图中的每个点  $(x, y)$  的值为图像中  $(x, y)$  点左上部分所有像素点的累加:

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y')$$

一旦计算出图像所对应的积分图,在任意尺度上的图像特征就能在常数时间内计算出来。如图 3 所示,在 1, 2, 3, 4 处点的积分图的值分别为 A,  $A + B$ ,  $A + C$ ,  $A + B + C + D$ , 而 D 的矩形特征可以通过  $4 + 1 - (2 + 3)$  计算出来,归纳起来就有如下公式计算任意尺度下的图像特征:

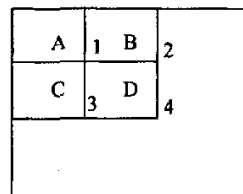


图 3 积分图像的概念

$$s(x, y) = s(x, y-1) + i(x, y)$$

$$ii(x, y) = ii(x-1, y) + s(x, y)$$

其中  $ii(x, y)$  为积分图中  $(x, y)$  点的值,  $s(x, y)$  为图

像中 $(x, y)$ 点以下列向量上的颜色值之和。 $i(x, y)$ 为 $(x, y)$ 点的像素值。这样,只要对输入图像遍历一次,也可以得到积分图,也可以在常数时间内得到任意尺度下的图像的特征。

### 3 瀑布型分类器

建立一个瀑布型分类器的主要思想是:对各种不同的输入窗口(对象)依照它们的复杂程度采用不同的处理方法以减少计算时间。如图4所示,只有输入窗口通过了瀑布型的所有层之后,才可以被分类做人脸区域。每层在训练时都只针对所有通过其上层分类器的样本进行训练。与前文提到的分层思想相似,这个框架的目标是尽快缩小候选区域,只有当上一级分类器返回真的时候才调用下一级更复杂的分类器进行判断。在这种灵活多变的结构下,更容易区分非人脸对象及类似结构的非人脸对象,并可以通过单一特征分类器滤除掉这些区域。

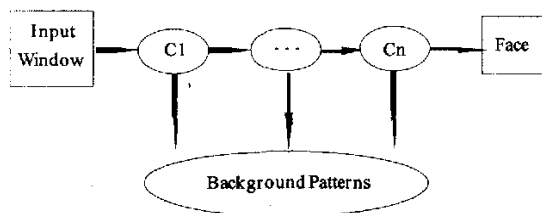


图4 瀑布型结构分类器

## 4 实验

### 4.1 训练数据集

为本次训练,找到了5000张 $24 \times 24$ 像素的人脸样本图,参见文献[5]。非人脸样本图在不同位置和范围内随机地出现,这些非人脸样本图来自于700多张不包含任何人脸信息,只是各种各样的石头、树木、建筑、风景、鲜花等,总之,大约有50万张非人脸样本图被用来训练我们的检测系统,为了训练基于分类器的 $36 \times 36$ 像素大小的窗口,使用了8000张 $24 \times 24$ 像素的人脸样本图,这些人脸样本图是从 $36 \times 36$ 像素窗口中随机转换成的 $24 \times 24$ 像素的窗口,正如前面一、二阶段所述,示例图如图1所示。

在训练 $24 \times 24$ 瀑布型分类器中,整个层分类器中使用了5000张相同的人脸样本图,其中,后来的层分类器中的非人脸样本图是不完全瀑布型在无脸图片集中收集的非正常图像。训练瀑布型分类器中的第一层分类器的无脸样本图是随机选取的,对每一层的分类器,在训练时都使用了总计5000张无脸样本图。在前两个阶段,确定的每层的检测率都为99.9%,其中,我

们控制了层的数量和每层的特征量的数量,以便在通过经验性的实验中兼顾检测速度和检测性能。经过前两个阶段所产生的信息在整个检测系统中就相当于过滤了一次一样,因此,比起文献[5]中的系统更显得灵活和容易。

### 4.2 系统结构

瀑布型分类器的第一阶段由3层组成,每层的特征数量分别为10,20,30;第二阶段由13层组成,前5层的特征数量分别为5,30,50,95,125,总之,第二阶段的13层总计有2475个特征,与文献[5]中系统中的6061个特征量相比,文中系统可以节省一半的训练时间。最后一个阶段的SVM分类器在第二阶段的最后一层使用了3200个特征量作为特征向量,并用LibSVM在RBF中心做了训练,其中, $C = 8, \gamma = 0.015625$ ,前两个阶段使用的层数量都是在经验性实验过程中总结出来的,它们兼顾了检测的速度和性能。

### 4.3 训练结果

我们在MIT+CMU正面人脸标准测试集<sup>[8]</sup>的环境中测试该系统,这个测试集包含了130张图像507张正面人脸。其性能在图5中显示,检测结果如图6所示。瀑布型分类器前两个阶段滤除掉了99.88%的无脸样本图,其中,第一阶段的分类器检测速度平均达到了58.73%以上,只有0.12%的复杂样本图进入最后一个阶段的SVM分类器中进行最后一轮检测。去除背景图所花费的时间大约占整个检测时间的38%,而利用SVM分类器分类的时间大约用去了62%。

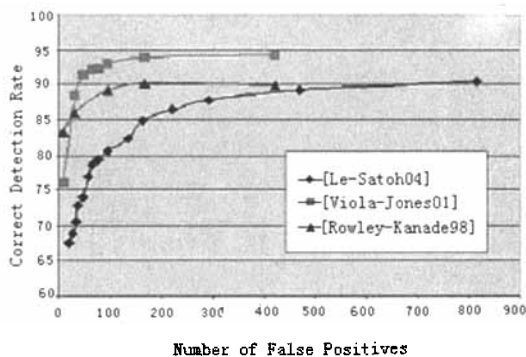


图5 人脸检测系统ROC曲线图



图6 本系统基于MIT+CMU测试集中图像检测结果  
(下转第138页)

三个像素内找到了毛刺生长点,则说明搜索到的这个端点不是图像中真正的端点,而是毛刺产生的“伪端点”。将“伪端点”和它与毛刺生长点之间的像素删除,即将毛刺删除。在全图范围内对所有搜索到的端点都进行上述处理,便可将细化图像上生长出来的毛刺都剔除。

## 2 实验结果与结论

文中的字符图像都经过了中值滤波、阈值分割等预处理,所采用的细化算法是 Zhang 细化算法。Zhang 快速并行细化算法具有速度快和保持细化后的连通性等优点,是一种实用的算法<sup>[9]</sup>。图 6 是利用上面方法对几个原来有毛刺的字符图像进行处理的结果。从结果中可以看出,原来图像中的毛刺都被剔除了,而字符真正的端点都被保留,达到了预期的效果。图 6(a)(b)(c)三幅图中左边的是带毛刺的字符,右边是去除毛刺后的字符。

综上所述,文中所采用的毛刺去除方法可以有效去除字符图像细化后所产生的毛刺,并且能够保证其它有用信息不被破坏,字符的整体结构特征得到恢复,效果稳定。为下一步提取字符的特征信息,提高字符的识别率提供了有力保证。

### 参考文献:

- [1] 李波,曾致远,周建中,等.车牌识别系统研究与实现[J].计算机技术与发展,2006,16(6):10-14.
- [2] 苑玮琦,张昱.纸币号码识别中分割方法的研究[J].计

(上接第 135 页)

## 5 总结

描述了一个基于 SVM 的加速的人脸检测系统,同时又保证了其相对的检测率。整个系统的提高是通过前两个阶段的分类器,这两个阶段的分类器加速了过滤背景图案的速度及计算时间,对检测潜在的人脸区域起到了很大的帮助作用。与选自 AdaBoost 学习机中的 Haar wavelet 特征的区别是:三层人脸检测系统将特征运用到了整个阶段的分类器中,并充分利用了它的高效展现和快速评估性能。前两个阶段分类器中的瀑布型结构能自适应各种不同的复杂性的输入对象。最后一个阶段中的非线性 SVM 分类器高速并有效地提供了检测的最佳结果。

### 参考文献:

- [1] Chellappa R, Wilson C L, Sirohey S. Human and machine

20050 20050

(a) 第一组字符

62829 62829

(b) 第二组字符

78725 78725

(c) 第三组字符

图 6 细化字符在毛刺去除前后的效果对比

算机测量与控制,2005,13(8):527-529.

- [3] 尚会超,陈幼平,余文勇,等.印刷字符在线检测的预处理算法及试验研究[J].计算机应用研究,2005(1):243-245.
- [4] 吴锐,赵巍,尹芳,等.特征融合及相似度判据在英文识别中的应用[J].计算机工程与应用,2005(16):55-57.
- [5] 吴谨,刘洋.具有抗噪稳健性的数字识别[J].武汉科技大学学报,2004,27(4):406-408.
- [6] 崔政,李壮.两种改进的模板匹配识别算法[J].计算机工程与设计,2006,27(6):1083-1085.
- [7] 王建雄,刘应龙.基于人工神经网络的数字识别系统的研究[J].计算机技术与发展,2006,16(5):26-30.
- [8] 谷口庆治.数字图像处理基础篇[M].北京:科学出版社,2002:106-109.
- [9] 章毓晋.图像工程(中册)-图像分析[M].第2版.北京:清华大学出版社,2002:220-223.

recognition of faces: A survey[J]. Proceedings of the IEEE, 1995, 83(5): 705-740.

- [2] 周杰,卢春雨,张长水,等.人脸自动识别方法综述[J].电子学报,2000,28(4):102-106.
- [3] 卢春雨.人脸自动识别若干问题研究与系统实现[D].北京:清华大学,1998.
- [4] Sung K, Poggio T. Example-based learning for view based human face detection[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1998, 20(1): 39-51.
- [5] Paul V, Michael J. Rapid Object Detection using a Boosted Cascade of Simple Features[C]//In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition. Kauai, Hawaii, USA:[s. n.], 2001.
- [6] Burges C J C. Tutorial on Support Vector Machines For Pattern Recognition[J]. Data Mining and Knowledge Discovery, 1998, 2(2): 121-167.
- [7] 梁路宏.人脸检测与跟踪研究[D].北京:清华大学,2001.
- [8] 张学工.关于统计学习理论与支持向量机[J].自动化学报,2000,26(1):32-42.