

## 精确分类的视角无关人脸检测方法与硬件加速体系结构

徐金波, 窦 勇

(国防科学技术大学并行与分布处理国家重点实验室 长沙 410073)  
(xujinbo@nudt.edu.cn)

**摘 要:** 为了对垂直于图像平面 $[-90^\circ, 90^\circ]$ 和图像平面内 $360^\circ$ 范围的人脸姿态进行快速准确的检测分类, 提出一种精确分类的视角无关人脸检测方法和可重构硬件体系结构. 设计了由多个检测节点组成的树形检测器框架, 并将多个姿态区间的分类问题采用向量式共享输出空间的方式统一起来, 提出一种两段式 Boosting 方法对检测节点进行训练; 挖掘了检测过程的时间空间并行性, 进而设计了一种高度并行的可重构硬件体系结构模板, 通过对体系结构模板进行动态配置, 实现了硬件代价、检测精度和速度的平衡. 实验结果表明, 与已有方法相比, 文中方法具有较高的精度与速度; 对于 $160 \times 120 \sim 800 \times 600$ 分辨率范围内的图像, 该方法在 FPGA 上的硬件实现与在 PC 上的软件实现的执行时间加速比为 14.68~20.86.

**关键词:** 视角无关人脸检测; Boosting; 可重构计算; FPGA  
**中图法分类号:** TP302; TP391

## A Fine-Classification Method and its Hardware Acceleration Architecture for Rotation Invariant Multi-View Face Detection

Xu Jinbo and Dou Yong

(National Laboratory for Parallel and Distributed Processing National University of Defense Technology, Changsha 410073)

**Abstract:** Aiming at detecting faces with all  $[-90^\circ, 90^\circ]$  degree rotation-out-of-plane and  $360^\circ$  rotation-in-plane pose changes efficiently and accurately, this work proposed a fine-classification method and an FPGA-based reconfigurable architecture for rotation invariant multi-view face detection. A coarse-to-fine tree-structured detector hierarchy composed of multiple detector nodes was designed. The proposed method deals with the multi-dimensional binary classification problems in a unified framework by means of a shared output space of multi-components vector. And a novel two-stage boosting method was proposed for training detector nodes. With the exploitation of both the spatial and temporal parallelism of the detection method, a highly parallel reconfigurable architecture template was designed. The reconfiguration of the architecture was evaluated for finding an appropriate tradeoff among the hardware implementation cost, the detection accuracy and speed. Experimental results on FPGA show that high accuracy and marvelous speed are achieved compared with previous related works. A speedup factor ranging from 14.68 to 20.86 for images of size of  $160 \times 120$  to  $800 \times 600$  is obtained compared with the conventional software solution on PC.

**Key words:** rotation invariant multi-view face detection; Boosting; reconfigurable computing; FPGA

对于人脸检测的研究已经取得了巨大的进展<sup>[1]</sup>, 典型的研究成果有文献[2-4]等, 但是这些方法计算代价较大. 2001 年, Viola 等<sup>[4]</sup>提出的基于提

升的层叠式框架取得了重大突破: 在保持较高检测精度的同时大大降低了计算代价, 提高了处理速度, 这得益于基于积分图的类 Haar 特征计算的高速度、

收稿日期: 2009-01-07; 修回日期: 2009-05-04. 基金项目: 国家自然科学基金重点项目(60633050, 60833004); 国家“八六三”高技术研究发展计划(2009AA01Z101). 徐金波(1981-), 男, 博士, 主要研究方向为高性能计算、并行体系结构等; 窦 勇(1966-), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究方向为高性能计算、并行体系结构等.

提升式强分类器的高精度以及层叠式的检测器结构的高效性。

尽管很多早期研究工作在正面人脸检测时具有较好的性能,但由于人脸姿态的复杂多样性,视角无关人脸检测(rotation invariant multi-view face detection, RIMVFD)仍然是一个具有挑战性的问题。RIMVFD 指的是垂直于图像平面 $[-90^\circ, 90^\circ]$ 范围(rotation-out-of-plane, ROP)以及图像平面内 $360^\circ$ 范围(rotation-in-plane, RIP)的各种人脸姿态的检测分类。

近年来,人们针对图像平面内视角无关人脸检测与平面外视角无关人脸检测(multi-view face detection, MVFD)提出了很多关于文献[4]方法的改进,大体分为以下 4 类:改进检测器结构<sup>[5-6]</sup>,改进强分类器设计<sup>[6-7]</sup>,改进弱分类器的训练<sup>[5,8]</sup>,以及改进特征的选择<sup>[9]</sup>。这些工作主要致力于提高检测精度,而且大部分工作仅仅是采用软件实现方案进行评价,并没有针对实时性要求较高的应用进行速度优化。

硬件加速技术能够较大幅度地实现算法到计算引擎的空间映射,从而提高算法处理速度,满足有高性能要求的专用应用领域对实时性的要求。本课题组已经对硬件加速技术中的设计空间探索方法<sup>[10]</sup>以及存储优化<sup>[11]</sup>等关键问题进行了深入的研究。为了满足不同的应用场景对人脸检测的需求,一种有效的解决方案是使用专用硬件对 RIMVFD 进行加速。目前虽然已有一些基于神经网络<sup>[12]</sup>和 AdaBoost<sup>[13]</sup>等方法的正面人脸检测硬件实现,但是对 RIMVFD 的硬件实现的研究较少。

本文提出了一种精确分类的视角无关人脸检测方法和一种可重构硬件体系结构。首先设计了一种由多个检测节点组成的树形检测器框架结构来实现对人脸目标的由粗到细的 RIP 和 ROP 姿态识别与分类。每个检测节点采用精确分类的 Boosting 方法进行训练;强分类器将不同姿态区间的多分类问题采用向量式共享输出空间的方式统一起来,并提出一种两段式 Boosting 方法对每个强分类器中的弱分类器进行训练。充分挖掘了本文方法的时间与空间并行性,设计了一种高度并行的可重构硬件体系结构模板,通过对该体系结构模板进行动态配置,实现了硬件实现代价、检测精度和速度之间的折中与平衡。

本文对组成强分类器的每个弱分类器提出了两段式 Boosting 训练方法:在第一阶段,根据每个样

本的类 Haar 特征值赋予该样本一个实值置信向量,向量的每一维置信值表示该样本属于每个姿态区间的概率(例如,第 1 维置信值表示该样本属于左侧面人脸的概率,第 2 维置信值表示该样本属于正面人脸的概率,第 3 维置信值表示该样本属于右侧面人脸的概率,等);在第二阶段,根据每个样本的置信向量进一步生成一个布尔值输出向量,该向量的第  $i$  个元素值使用布尔值指明了该样本是否被认为属于第  $i$  个姿态区间,在此过程中,布尔值输出向量的每个元素的计算都需要第一阶段输出的置信向量所有元素的参与,第一阶段置信向量各元素与第二阶段布尔值输出向量各元素之间存在着全互联的映射关系。也就是说,要判断一个样本是否属于某个姿态区间,不仅要看第一阶段中该样本属于相应姿态区间的概率有多大,还要看该样本不属于其他姿态区间的概率有多大。由于这种 2 个训练阶段之间的全互联映射关系考虑了不同的姿态区间对应的置信值的相关性,因此有助于对样本所属的姿态区间进行更精确的分类。在文献[6]中,由于判断一个样本是否属于某个姿态区间时只使用该样本对应当前姿态区间的置信值,而没有考虑该样本不属于其他姿态区间的概率,因此某些姿态区间可能被不必要地判断为该人脸样本所处的候选区间,从而增加了分类过程的计算复杂性。

## 1 RIMVFD 方法

### 1.1 基于 AdaBoost 的人脸检测框架

由于基于 AdaBoost 的人脸检测具有较高的速度和精度,因此得到了人们较广泛的关注,AdaBoost 的基本思想是若干个“弱分类器”组合成一个“强分类器”。Viola 等<sup>[14]</sup>提出了将类 Haar 特征应用于 AdaBoost 人脸检测的方法,类 Haar 特征可以使用积分图快速地进行计算。在 Viola 的层叠式框架中,一组特征基于 AdaBoost 构成一个分类层次,每个分类层次的输出决定了被检测的图像区域是否包含一个人脸目标。当图像中所有基本大小的区域被处理完后,将特征尺寸按照缩放系数逐级进行放大,从而实现对图像中所有尺寸区域的人脸检测。图 1 所示为基于 AdaBoost 进行人脸检测的示意图,详细的方法描述见文献[14]。本文提出的 RIMVFD 方法与图 1 所示框架的主要区别在于检测器部分,而积分图的计算与图像缩放模块是相同的。下面主要介绍 RIMVFD 方法的检测器部分的设计。

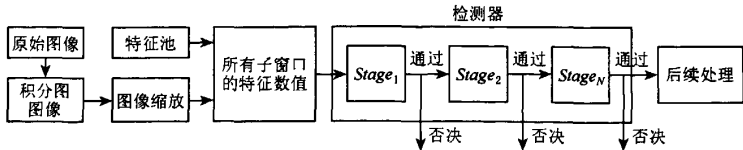


图 1 基于 AdaBoost 的人脸检测框架

1.2 树形 RIMVFD 检测器框架结构

对于 RIMVFD,除了人脸/非人脸的分类外,还需要对人脸的姿态进行识别与分类<sup>[5-6]</sup>,因此 RIMVFD 属于多类分类问题.文献[6]提出的树形结构能够较好地平衡人脸/非人脸分类任务以及姿态识别任务,从而提高人脸检测的速度和精度.但由于受到通用处理器的并行处理能力的限制,文献[6]中的树形结构在 ROP 平面内只是把人脸姿态划分

为 5 个区间来控制计算代价.

如图 2 所示,本文设计了一种细粒度的树形检测器框架结构,通过配置较多的检测节点并借助硬件加速的帮助,来获得更高的检测速度和精度.该结构覆盖了 $[-90^{\circ}, 90^{\circ}]$ ROP 和  $360^{\circ}$ RIP 范围内的人脸姿态变化,并且在 ROP 和 RIP 平面内的区间粒度分别为  $20^{\circ}$ 和  $30^{\circ}$ ;通过采用由粗到细的策略,将整个人脸姿态空间逐层划分为越来越小的子空间.

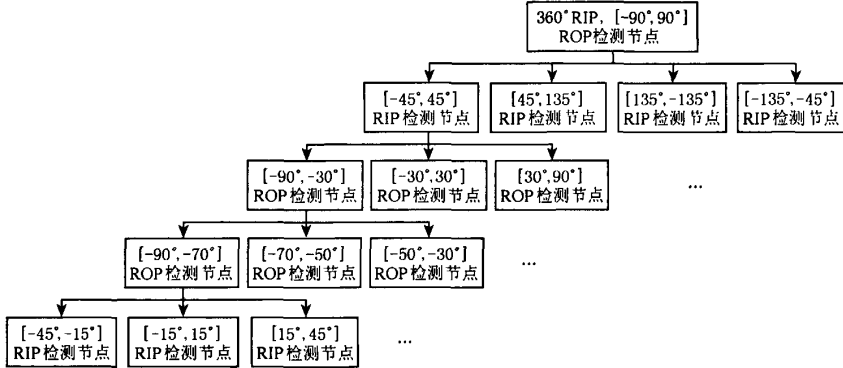


图 2 细粒度的树形检测器框架结构示意图

一个人脸样本是否被认为属于某个姿态区间是通过该姿态区间对应的检测节点的布尔类型输出值  $g(x)$  决定的.如图 2 中具有同一个父节点的、所有检测节点的  $g(x)$  使用属于该父节点的一组样本集进行训练,这些  $g(x)$  组成一个决策向量  $G(x) = (g_1(x), \dots, g_k(x))$ ,其中  $k$  为该父节点的子检测节点的个数.当人脸样本经过逐级处理达到某个或某些叶节点时,该样本所处的 RIP 和 ROP 姿态区间即可被判断出来.

1.3 精确分类的两段式 Boosting 方法

如前所述,RIMVFD 是一个多类分类问题,以图 2 中 $[-45^{\circ}, 45^{\circ}]$  RIP 检测节点为例,通过该检测器的处理后,人脸样本将被送入 $[-90^{\circ}, -30^{\circ}]$ ROP,  $[-30^{\circ}, 30^{\circ}]$ ROP 和  $[30^{\circ}, 90^{\circ}]$ ROP 3 个检测节点,以进一步判断该样本属于这 3 类姿态区间中的哪一类,此处  $G(x)$  为一个三维向量,如  $G(x) = (1, 0, 1)$

表示该样本可能属于 $[-90^{\circ}, -30^{\circ}]$ ROP 姿态范围,也可能属于 $[30^{\circ}, 90^{\circ}]$ ROP 姿态范围.

人们提出了很多基于基本的 AdaBoost 方法的变种来实现 MVFD<sup>[6-7]</sup>.文献[6]中提出的 Vector Boosting 方法将不同姿态区间的多类分类问题采用向量式共享输出空间的方式统一起来,每一类分类问题在输出空间中都有自己的“兴趣方向”,称为“投影向量”.首先采用 Real AdaBoost 的扩展形式为当前样本对应每个姿态区间计算一个置信值;然后把由这些置信值组成的置信向量与一个阈值向量  $B$  进行比较,得到一个布尔值的输出向量,每个输出值表示当前样本是否属于某个姿态区间.由于这种方法在判断样本是否属于某个姿态区间时只使用该样本对应当前姿态区间的置信值,而没有考虑该样本对应其他姿态区间的置信值,因此相当于将姿态区间的多分类问题分解成每个姿态区间的 2 分类问题.

使用这种方法,某个人脸样本可能会被认为属于多个姿态区间,而事实上,大多数人脸样本一般只应该属于某一个姿态区间。例如,本节第 1 段提到的例子中的样本可能只属于  $[-90^\circ, -30^\circ]$  ROP 姿态区间,而计算出的  $G(x)$  却可能为  $(1, 0, 1)$ , 导致了分类准确性的降低。

为了提高分类准确性,本文提出了一种精确分类的 Boosting 方法来训练每个检测节点,目标是减少样本被错误分类的情况。例如,对  $G(x) = (1, 0, 1)$  采用本文提出的方法可以精确到  $G(x) = (1, 0, 0)$ 。实现该目标的技术途径也是本文在检测算法上的主要创新之处:对组成每个检测节点中强分类器的每个弱分类器提出了两段式 Boosting 训练方法。具有同一个父节点的所有  $k$  个检测节点使用共同的一组样本集进行弱分类器训练,这里使用  $h_i^*(x)$  表示由  $k$  个使用相同训练样本集的弱分类器组成的弱分类器向量。 $h_i^*(x)$  的第一阶段  $h_i'$ :  $\mathcal{X} \rightarrow \mathbb{R}^k$  基于特征分片的思想(其中  $\mathcal{X}$  为样本空间,  $\mathbb{R}^k$  为  $k$  维实值置信空间)。该阶段的分类函数根据每个样本的类 Haar 特征值赋予该样本一个实值的  $k$  维置信向量  $\{x_i | i=1, \dots, k\}$ , 每个向量元素表示该样本属于该维向量元素所对应的姿态区间的概率(例如,第 1, 2, 3 维置信值可以分别表示该样本属于左侧面、正面和右侧面人脸的概率)。 $h_i^*(x)$  的第二个阶段  $h_i''$ :  $\mathbb{R}^k \rightarrow \{-1 | +1\}^k$  基于超矩形的思想,该阶段的分类函数根据每个样本的置信向量进一步生成一个  $k$  维的布尔值输出向量,其第  $i$  个元素的值指明了该样本是否被认为属于第  $i$  个姿态区间。在这一过程中,任意一维布尔类型输出值都依赖于第一阶段中所有  $k$  维置信值,而不是只依赖于对应的某一维置信值。 $h_i'$ :  $\mathcal{X} \rightarrow \mathbb{R}^k$  和  $h_i''$ :  $\mathbb{R}^k \rightarrow \{-1 | +1\}^k$  连接起来形成了一个完整的弱分类器  $h_i^*(x)$ :  $\mathcal{X} \rightarrow \{-1 | +1\}^k$ , 该弱分类器的第  $i$  维输出反映了它对于某个样本是否属于第  $i$  个姿态区间的判断。

对弱分类器的训练过程不断地进行迭代,直到得到若干个分类误差最小的弱分类器,所有训练出的弱分类器通过线性组合得到一个  $k$  维的共享输出空间  $H(x)$ 。通过对该共享输出空间施加不同的投影向量来处理  $n$  个姿态区间的分类问题,一个由  $n$  个  $k$  维投影向量组成的  $n \times k$  矩阵  $A$  用来将  $k$  维的输出结果转换到  $n$  维的置信空间  $F(x)$  中,置信空间中的每一维对应到分类问题中某一个姿态区间的检

测。最终,强分类器的输出为  $F(x)$  的符号。

现仍然继续以本节第 1 段的例子为例,对  $[-90^\circ, -30^\circ]$  ROP,  $[-30^\circ, 30^\circ]$  ROP 和  $[30^\circ, 90^\circ]$  ROP 3 个检测节点的分类器使用相同的训练样本集进行训练,此处共享输出空间的维数  $k$  为 3。如果需要处理  $[-90^\circ, -30^\circ]$  ROP 和  $[30^\circ, 90^\circ]$  ROP 2 个姿态区间的分类问题,那么  $n=2$ ,  $n \times k = 2 \times 3$  矩阵  $A$  为  $((1, 0, 0), (0, 0, 1))$ 。通过投影向量  $(1, 0, 0)$  可以从共享输出空间中取出第 1 维的输出值,来进行  $[-90^\circ, -30^\circ]$  ROP 姿态区间的检测;类似地,通过投影向量  $(0, 0, 1)$  可以从共享输出空间中取出第 3 维的输出值,进行  $[30^\circ, 90^\circ]$  ROP 姿态区间的检测。

下面给出了这种方法的算法描述。定义训练集中的  $m$  个样本为  $S = \{(x_1, v_1, y_1), \dots, (x_m, v_m, y_m)\}$ , 其中,  $x_i$  属于样本空间  $\mathcal{X}$ ;  $v_i$  属于一个有限的  $k$  维投影向量集合  $\Omega$ , 用来表示该训练样本所对应的姿态区间;而  $y_i = \pm 1$ , 表示该样本是否属于  $v_i$  所示的姿态区间。对于每个样本,弱分类器的判断结果与样本  $x_i$  的实际情况之间的差距用  $y_i \cdot (v_i \cdot h_i^*(x_i))$  表示。例如,  $y_i = 1$  表示该样本  $x_i$  属于  $v_i = (1, 0, 0)$  所对应的  $[-90^\circ, -30^\circ]$  ROP 姿态区间,而通过训练得到的分类器  $h_i^*(x_i)$  对样本  $x_i$  的判断结果却为  $-1$ , 这表示分类器认为该样本不属于  $[-90^\circ, -30^\circ]$  ROP 姿态区间,即分类器判断结果与实际情况不符,分类器判断错误。此时,  $y_i \cdot (v_i \cdot h_i^*(x_i)) = -1$ ; 否则,如果分类器对样本  $x_i$  的判断结果为 1, 即分类器判断结果与实际情况相符,分类器判断正确。此时  $y_i \cdot (v_i \cdot h_i^*(x_i)) = 1$ ,  $y_i \cdot (v_i \cdot h_i^*(x_i))$  将被用于对样本权重进行更新。由于投影向量  $v_i$  的引入,弱分类器输出向量中与当前姿态区间正交的元素将不会对弱分类器的训练过程产生任何影响。

两段式 Boosting 算法描述如下:

对于一个  $n$  个姿态区间的分类问题,给定

1) 投影向量集合  $\Omega = \{\omega_1, \dots, \omega_n\}$ ,  $\omega_i \in \mathbb{R}^k$ ;

2) 样本集合  $S = \{(x_1, v_1, y_1), \dots, (x_m, v_m, y_m)\}$ , 其中  $x \in \mathcal{X}$ ,  $v \in \Omega$ ,  $y = \pm 1$ 。

Step1. 初始化所有  $m$  个样本的权重  $D_1(i) = 1/m$ ,  $i = 1, \dots, m$ 。

Step2. FOR  $t = 1, \dots, T$

使用当前样本权重分布来训练弱分类器第一阶段分类函数  $h_t'$ :  $\mathcal{X} \rightarrow \mathbb{R}^k$ ;

使用当前样本权重分布来训练弱分类器第二阶段分类函数  $h_t''$ :  $\mathbb{R}^k \rightarrow \{-1 | +1\}^k$ ;

第一阶段和第二阶段的分类函数通过全互联方式连接起来形成当前迭代的弱分类器  $h_i^*(x)$ :  $\mathbf{x} \rightarrow \{-1|+1\}^k$ ;

计算  $h_i^*(x)$  的加权误差  $\epsilon_i = \sum_{i=1}^m D_i(i) \cdot I(y_i \neq v_i \cdot h_i^*(x_i))$ ;

计算系数  $\alpha_i = \frac{1}{2} \log \left( \frac{1-\epsilon_i}{\epsilon_i} \right)$ ;

对样本权重进行更新

$D_{i+1}(i) = \frac{D_i(i) \cdot \exp(-\alpha_i \cdot y_i \cdot (v_i \cdot h_i^*(x_i)))}{Z_i^*}$ , 其中  $Z_i^*$  是归一化因子。

Step3. IF ( $\epsilon_i = 0$  OR  $\epsilon_i \geq 1/2$  AND set  $T = t-1$ ), 停止。

Step4. 共享输出空间为  $\mathbf{H}(x) = \sum_{i=1}^T \alpha_i h_i^*(x)$ 。

Step5. 置信空间为  $\mathbf{F}(x) = \mathbf{A}\mathbf{H}(x)$ , 其中转换矩阵  $\mathbf{A} = \{\omega_1, \dots, \omega_n\}^T$ 。

Step6. 最终的强分类器为  $\mathbf{G}(x) = \text{sgn}(\mathbf{F}(x))$ 。

#### 1.4 弱分类器的第一阶段

弱分类器的第一阶段基于特征分片的思想<sup>[5]</sup>, 它比基于阈值的分类函数更加有效, 而且可以使用查找表(look up table, LUT)结构实现高效的分类。基于特征分片的分类函数由特征空间的划分和每个划分区间(用  $\text{bin}$  表示)的置信常数  $c$  2 个部分组成。对于第  $i$  维姿态区间弱分类器的训练, 假定每个训练样本的特征值  $f_{\text{Haar}}$  已经作了归一化处理, 特征空间被平均划分为  $p$  个子区间:  $\text{bin}_j = [(j-1)/p, j/p]$ ,  $1 \leq j \leq p$ , 其中每个子区间对应样本空间  $\mathbf{x}$  中的一部分样本, 那么  $h_i'$  定义为  $f_{\text{Haar}}(x) \in \text{bin}_j \Rightarrow h_i'(x) = c_j$ , 即当该样本的类 Haar 特征值属于其中的第  $j$  个划分区间时, 弱分类器输出的置信值为对应该划分区间的置信常数  $c_j$ , 其中  $c_j$  可以使用一种合适的最优化算法(如 Newton-Step 方法)进行训练。综上,  $h_i'$  可以表示为

$$h_i'(x) = \sum_{j=1}^p c_j^* B_p^j(f_{\text{Haar}}(x)), \quad (1)$$

其中,  $B_p^j(u) = \begin{cases} 1, & u \in [(j-1)/p, j/p] \\ 0, & u \notin [(j-1)/p, j/p] \end{cases}, j=1, \dots, p;$

$c_j^* = \{c_{j(i)} \mid i=1, \dots, k\}; h_i' = \{h_{i'(i)}' \mid i=1, \dots, k\}$ 。

#### 1.5 弱分类器的第二阶段

弱分类器的第二阶段通过对所有样本的置信向量分布情况进行学习, 以构造一个分类函数对不同的姿态区间进行更精确的划分。弱分类器的第二阶段基于超矩形的思想, 它比基于单个阈值的分类函数能产生更好的结果, 而且比较容易进行并行优化。定义一般的超矩形为由  $2k$  个阈值和一个布尔值  $y_H$  (即  $y_H \in \{-1, +1\}$ ) 组成的集合  $\mathbf{H}: \mathbf{H} = \{\theta_1^l, \theta_1^u, \theta_2^l,$

$\theta_2^u, \dots, \theta_k^l, \theta_k^u, y_H\}$ , 其中  $\theta_i^l$  和  $\theta_i^u$  分别是第  $i$  维空间的下限和上限。则分类函数定义为

$$h_H(x) = \begin{cases} y_H \Leftrightarrow \prod_{i=1}^k ((x_i > \theta_i^l) \text{ and } (x_i < \theta_i^u)) \\ -y_H, & \text{otherwise} \end{cases} \quad (2)$$

其中  $y_H = 1$  表示当某样本的置信值  $x_i$  满足式(2)第1个式子所示的条件时, 弱分类器认为该样本属于该分类器对应的姿态区间;  $y_H = -1$  表示当某样本  $\mathbf{x}$  的置信值  $x_i$  满足式(2)第2个式子所示的条件时, 弱分类器认为该样本不属于该分类器对应的姿态区间。

弱分类器第二阶段训练过程的目标是通过一个训练样本集  $\mathbf{S}$  生成一个超矩形集合  $\mathbf{S}_H$ , 本文使用文献[15]方法进行训练, 其基本思想是在每个样本  $\{x_i, y_i\} \in \mathbf{S}$  的周围构建一个超矩形  $H(x_i)$ , 使得该矩形不会包含其他姿态区间内的样本。在本文方法中,  $x_i$  是弱分类器第一阶段训练过程所用样本的置信向量,  $y_i = \pm 1$ 。通过使用所有样本对应的超矩形  $H(x_i)$  进行训练, 最终找出使得训练样本的分类误差最小的超矩形集合  $\mathbf{S}_H$  对应的  $\theta^l$  和  $\theta^u$ , 即可根据式(2)确定弱分类器的分类函数。训练样本的分类误差是指, 当前训练出的弱分类器对样本属于哪个姿态区间的判断是否符合该样本所属姿态区间的实际情况。

## 2 硬件体系结构的设计与实现

### 2.1 总体结构设计

首先挖掘 RIMVFD 方法的时间与空间并行性。图像获取、积分图计算与人脸分类3个阶段以数据驱动的工作方式形成一个粗粒度的流水线: 图像获取从数据源获得图像数据, 由积分图计算过程生成该图像的积分图, 人脸分类阶段根据积分图进行特征计算与分类。图像缩放模块和人脸分类阶段可以并行工作, 当人脸分类模块对当前缩放层次的图像进行处理时, 图像缩放模块同时将图像缩放到其他的缩放层次。本文中是对图像而不是特征子窗口进行缩放, 保证了所有特征子窗口的大小保持不变, 有利于保持较高的数据局部性, 以降低在缓存较小的专用硬件中实现时由于大量数据访问失效导致的性能下降。在人脸分类阶段, 图2所示树形检测器结构中的不同的层次以流水线的形式对不同的子窗口进行处理, 而同一层次中不同检测节点可以并行执行。通常, 不同层次中检测节点的数目是不同的,



层次越低,检测节点越多.

根据以上的并行特性分析,本文设计了图 3 所示的 RIMVFD 总体结构,包括积分图计算模块、人脸分类模块和图像缩放模块.积分图计算模块的结果送入人脸分类模块和图像缩放模块,人脸分类模块还需要用到图像缩放模块生成的不同缩放层次的图像.每个检测节点包括一个特征计算部件和一个分类函数实现模块,所有模块生成的中间结果通过一个多端口的存储控制器保存在主存储器中.已经有很多关于积分图的计算和图像的缩放的研究工作<sup>[13,16]</sup>,本文重点研究人脸分类模块的设计.

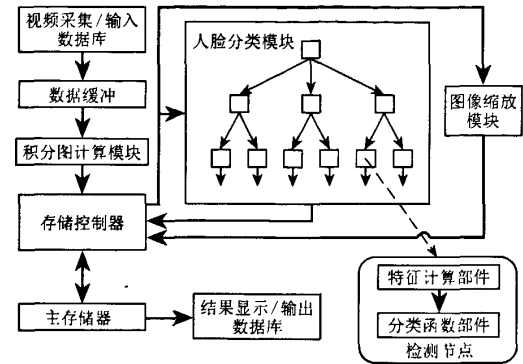


图 3 RIMVFD 系统的总体结构

本文通过动态重构的策略,在不明显影响检测速度的前提下,将图 2 所示检测器中检测节点的数目降低了约 75%.由于图 2 中以  $[-45^\circ, 45^\circ]$ ,  $[45^\circ, 135^\circ]$ ,  $[135^\circ, -135^\circ]$  和  $[-135^\circ, -45^\circ]$  4 个 RIP 姿态区间检测节点为根节点的分支可以重用同样的结构,它们的唯一区别仅在于进行特征计算时所需要数据的存储器访问地址,因此可以先将  $[-45^\circ, 45^\circ]$  RIP 姿态区间检测分支和图 2 中整个检测器的根节点配置在系统中.如果一个子窗口被根节点否决,则认为该子窗口不存在人脸目标;否则,该子窗

口被送入  $[-45^\circ, 45^\circ]$  RIP 检测节点进行处理,直到该子窗口的检测结果从该检测分支的某个或某些叶节点输出.被  $[-45^\circ, 45^\circ]$  RIP 检测节点否决的待检测子窗口的坐标信息将缓存起来.当图像中所有子窗口被  $[-45^\circ, 45^\circ]$  RIP 检测节点及其分支处理完毕后,对配置在系统中的所有检测节点的特征计算部件进行更新,使得相应的特征可以旋转  $90^\circ$ ,  $180^\circ$  和  $270^\circ$ .缓存起来的子窗口将被重新由更新过的检测器进行处理,只有那些被前一个  $90^\circ$  RIP 检测节点及其分支否决的子窗口才会送入当前  $90^\circ$  RIP 检测节点及其分支进行进一步处理.因为现实生活中大多数人脸目标集中在  $[-45^\circ, 45^\circ]$  RIP 姿态区间内,所以系统中被  $[-45^\circ, 45^\circ]$  RIP 检测节点否决的子窗口的数量很少.采用这种动态重构策略后,尽管图 2 中检测节点的数目由 161 减少到 41,但是检测速度和精度将几乎不会受到影响.

2.2 检测节点间的组织与通信

树形检测器框架中不同的检测节点具有相似的结构.每个节点的输入数据由待检测子窗口的特征数值和用来指示当前节点是否需要被激活的使能信号组成,输出数据是用来指示待检测子窗口是否属于当前人脸姿态区间的一个布尔值信号.一个检测节点的所有直接子节点使用由同一个特征计算部件生成的特征数值集合.对于非叶节点,输出结果被送入其所有子节点作为这些子节点的使能信号;否则,叶节点的输出结果即为当前子窗口的检测结果.在每个节点中,采用层叠式结构<sup>[4]</sup>组织强分类器,这种结构的优点是早期阶段中较简单的强分类器可以有效地过滤掉大部分不包含兴趣人脸的子窗口.前一个强分类器的输出作为后一个强分类器的使能信号,不同层次的强分类器以流水线的方式对子窗口进行处理.使用相邻的两层检测节点之间的组织与通信如图 4 所示.

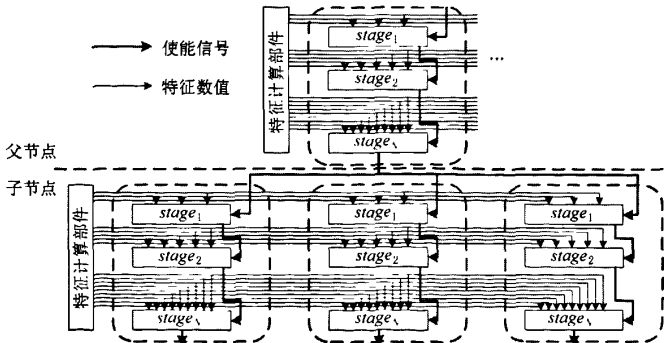


图 4 检测节点间的组织与通信

### 2.3 强分类器的并行实现

不同的强分类器具有相似的结构,只是分类函数的参数以及弱分类器的数量有所不同。每个强分类器的输出为若干乘积的累加和的符号,其中每个乘积由一个常数  $\alpha_i$  和由  $h_i^*$  输出决定的一个  $-1$  或  $+1$  的数值组成,因此这些乘积的累加和实际上是若干  $\alpha_i$  的加减法操作。本文将  $\alpha_i$  的加减法操作分组,并将中间结果事先编码到 LUT 部件中;弱分类器  $h_i^*$  的输出用来作为 LUT 部件的访问地址。LUT 部件中每个地址所对应的加减法操作越多,所需要的在线加减法操作就越少。图 5 所示为使用 16 地址 LUT 部件实现强分类器的结构。

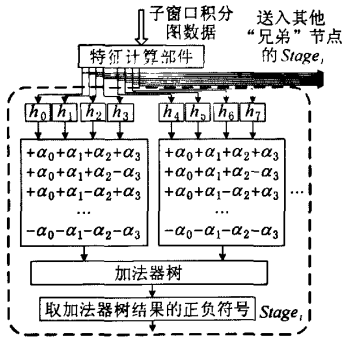


图 5 强分类器的结构

### 2.4 两段式弱分类器的并行实现

第一阶段分类函数  $h_i^*: \mathbf{x} \rightarrow \mathbb{R}^k$  具有式(1)所示的形式,其中  $c_j^*$  的训练是离线的;第二阶段分类函数  $h_i'': \mathbb{R}^k \rightarrow \{-1|+1\}^k$  对于  $k$  维分类问题的每一维都具有式(2)所示的形式。对于每个强分类器中的第  $i$  个弱分类器,待检测子窗口的特征数值首先被送入所有  $k$  个  $h_i'$ ,其中每个  $h_i'$  对应一个姿态区间,所有  $k$  个输出将以全互联的映射关系送入所有  $k$  个  $h_i''$ 。图 6 所示为弱分类器的工作流程。

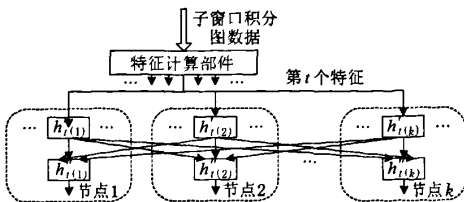


图 6 弱分类器的工作流程

不同的  $h_i'$  具有相似的结构。基于特征分片的分类函数可以采用图 7 a 所示的 LUT 部件进行并行实现,LUT 中的每个元素对应一个特征划分区间(用  $bin$  表示)的置信常数。输入数据为归一化的特

征数值,它们作为 LUT 的访问地址。如果特征数值位于所有  $p$  个特征区间的第  $j$  个,它对应的  $c_j$  即为输出的结果。不同的  $h_i''$  也具有相似的结构。基于超矩形的分类函数可以采用图 7 b 所示的硬件结构实现并行处理,该部件只包括一些比较部件和逻辑运算部件。

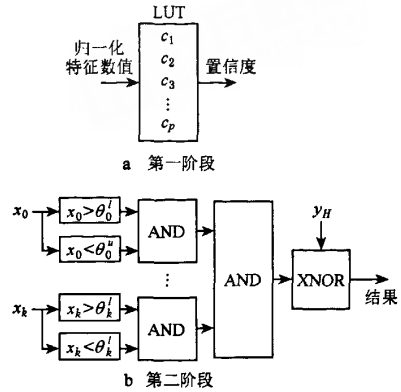


图 7 两段式弱分类器的硬件结构

### 2.5 树形检测器结构的动态配置

本文通过对树形检测器的硬件体系结构模板进行动态配置来实现硬件实现代价、检测精度和速度之间的折中与平衡。整个树形检测器结构由若干个检测节点组成,每个检测节点由多个以层叠式结构组织的强分类器组成,而每个强分类器又由若干个弱分类器构成。通常,在以层叠式结构组织的强分类器中,早期阶段中较简单的强分类器虽然包含较少的弱分类器,却可以有效地过滤掉大部分不包含感兴趣人脸的子窗口;而后期阶段中较复杂的强分类器包含较多的弱分类器,用来对某些较难分类的子窗口进行进一步检测分类。如果在检测器结构中配置较多的强分类器,而每个强分类器配置较多的弱分类器,那么系统可以实现较高的 RIMVFD 检测效果,但这是以更多的硬件资源为代价的,详见第 3.5 节实验分析。

## 3 实验结果及分析

下面通过实验来对本文提出的 RIMVFD 方法和硬件结构进行评价分析。我们收集了超过 85 000 个人脸样本,并将它们标准化到  $24 \times 24$  像素大小,这些样本覆盖了所有的  $[-90^\circ, 90^\circ]$ ROP 和  $[-45^\circ, 45^\circ]$ RIP 范围内的人脸姿态。将样本空间由粗到细地逐渐划分为姿态覆盖范围越来越小的样本子空间,然后采用第 1.3~第 1.5 节的训练方法对这些

样本进行训练,得到最终的树形检测器结构,该检测器由 5 层共 41 个检测节点组成.如第 2.1 节所述,对 $[45^{\circ},135^{\circ}]$ , $[135^{\circ},-135^{\circ}]$ 和 $[-135^{\circ},-45^{\circ}]$  RIP

姿态区间的检测分类可以通过对 $[-45^{\circ},45^{\circ}]$  RIP 姿态区间检测节点及其分支进行 $90^{\circ},180^{\circ}$ 和 $270^{\circ}$ 顺时针旋转来实现.图 8 所示为一些训练样本.



图 8 一些训练样本

3.1 测试平台与测试数据

我们在 FPGA 上根据本文方法和硬件体系结构设计了一个原型系统,其主要包括一个 FPGA 芯片和一个 SDRAM 模块,通过 USB 接口与宿主计算机相连.FPGA 芯片采用 Altera Stratix II EP2S130F1020C5,其包括 106 032 个 ALUT 和 6 747 840 位的片上存储器;SDRAM 模块容量为 1 GB,用来作为片外主存储器;宿主计算机担当视频采集设备和显示/存储设备的角色.系统设计采用 Verilog HDL 进行编码,采用 Mentor Graphics ModelSim 进行仿真,并使用 Quartus II 进行综合.

本文使用 CMU 人脸轮廓测试集对系统进行测试,该测试集包含 208 幅图像共 441 张人脸.由于 CMU 人脸轮廓测试集不能很好地评价平面内视角无关的人脸检测,我们从不同的数据源收集了包含所有 $[-90^{\circ},90^{\circ}]$ ROP 姿态区间和 $360^{\circ}$  RIP 范围的 213 幅图像共 682 张人脸.为了评价系统的正面人脸检测效果,我们还使用 CMU+MIT 正面人脸测试集对本文系统进行测试,该测试集包含 130 幅图像共 507 张人脸.图 9 所示为本文系统产生的一些人脸检测结果.



图 9 本文系统的一些人脸检测结果

3.2 资源使用情况

我们使用 Quartus II 综合对本文系统的资源使用情况进行了分析.子窗口的大小 $x_0 \times y_0$ 为 $24 \times 24$ , $x$ 方向和 $y$ 方向扫描的步幅 $S_x$ 和 $S_y$ 均设为 1 个像素,缩放系数 $r$ 为 1.25;每个检测节点中的强分类器的个数以及每个强分类器中弱分类器的个数通过在 ROC(receiver operator characteristic)曲线上选择合适的点来确定.最终,我们在检测器结构中采用了 3 820 个弱分类器.基于综合结果,系统的时钟频率达到 98 MHz.表 1 所示为本文系统各个模块的资源使用情况,可以看出人脸分类模块占用了大部

分的 ALUT 资源,原因在于本文在检测器中配置了大量的弱分类器.

表 1 系统各个模块的资源使用情况

模块	面积/ALUT 个数	片上存储器/位
视频采集/输入数据库接口	238	67 239
结果显示/输出数据库接口	322	8 325
主存储器控制器	819	19 200
积分计算模块	657	48 921
人脸分类模块	102 512	69 038
图像缩放模块	579	32 973
总计	105 127	245 696



3.3 速度比较

我们将本文系统与 PC(P4 2.8 GHz CPU,1 GB 内存)上的软件实现方案的执行速度进行了比较. 软件实现方案基于 OpenCV 实现,在 Windows XP 操作系统下采用 Visual C++ 语言编写而成. 表 2 所示为使用软硬件实现方案检测一幅图像中的视角无人脸目标时的执行延迟的比较以及片外存储器的带宽需求情况,本文采用了各种不同大小的测试图像. 可以看出,尽管本文系统的时钟频率只有 98 MHz,与 PC 的 2.8 GHz 相比明显要慢,但是硬件方案的执行速度仍然比软件方案要快很多. 对于 160×120~800×600 的测试图像,硬件方案执行时间加速比达

到了 14.68~20.86;对于 320×240 的视频序列,只要视频采集设备或输入数据库能够足够快地生成原始图像,系统的处理速度可以达到 186 帧/s,比软件实现方案快 17.52 倍. 即使对于 800×600 的视频序列,处理速度也可以达到 35 帧/s. 这种执行速度比其在 PC 或 FPGA 上实现的相关工作<sup>[5-6,16]</sup>也要快很多,原因在于我们充分挖掘了整个系统尤其是检测器结构的时间与空间并行性,从而实现了高度并行和流水化的硬件体系结构. 本文系统的可用片外存储器带宽约为 200 MB/s,从表 2 中可以看出,这可以满足系统处理各种不同大小图像时的带宽需求.

表 2 软硬件执行时间比较与片外存储器带宽需求

图像大小	执行延迟/ms		加速比	片外存储器带宽需求/MB/s
	硬件(98 MHz)	软件(2.8 GHz)		
160×120	1.58	23.2	14.68	99.8
192×144	2.06	33.6	16.31	110.7
320×240	5.36	93.9	17.52	158.7
384×288	7.27	135.4	18.62	168.5
480×320	9.71	188.2	19.38	175.4
640×480	18.42	376.6	20.45	185.1
768×576	26.08	542.5	20.8	188.3
800×600	28.22	588.7	20.86	188.9

3.4 精度比较

1) CMU+MIT 测试集上的正面人脸检测

由于某些应用对正面人脸检测具有一定的需求,首先测试本文方法在 CMU+MIT 测试集上的正面人脸检测,并与 ANN (artificial neural network)方法<sup>[2]</sup>、层叠式 AdaBoost 方法<sup>[4]</sup>和瀑布型人脸检测方法<sup>[5]</sup>进行了比较. 图 10 所示为上述 4 种方法在 CMU+MIT 测试集上的 ROC 曲线,可以

看出,本文方法具有较高的检测率.

2) CMU 人脸轮廓测试集上的 MVFD 测试

对于 MVFD,我们使用 CMU 人脸轮廓测试集将本文方法与 Bayesian 决策方法<sup>[3]</sup>、瀑布型人脸检测方法<sup>[5]</sup>和 WFS (width-first-search) 树方法<sup>[6]</sup>进行了比较. 图 11 所示为上述 4 种方法对应的 ROC 曲线.

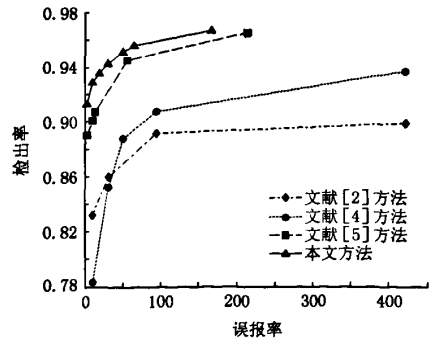


图 10 CMU+MIT 测试集上的 ROC 曲线

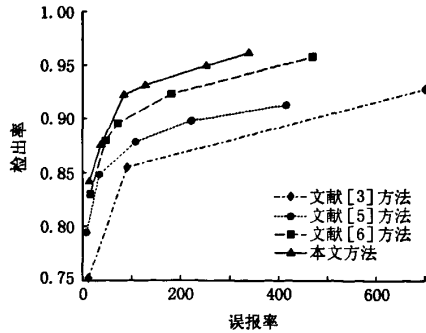


图 11 CMU 人脸轮廓测试集上的 ROC 曲线

3) 本文收集的测试集上的 RIMVFD 测试

对于 RIMVFD,由于目前没有标准的测试集,

本文从不同的数据源收集了包含所有  $[-90^\circ, 90^\circ]$  ROP 和  $360^\circ$  RIP 范围的 213 幅图像共 682 张人脸对本文方法进行测试. 瀑布型人脸检测方法<sup>[5]</sup>和 WFS 树方法<sup>[6]</sup>虽然也针对 RIMVFD 进行了一些测试, 但其中没有给出 ROC 曲线. 图 12 所示为本文方法在本文收集的测试集上的 ROC 曲线.

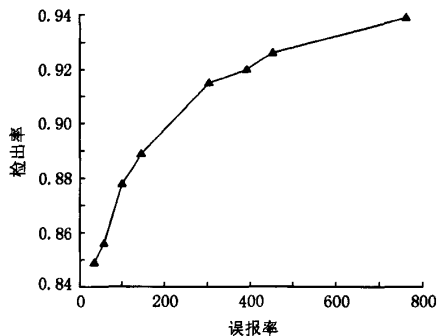


图 12 本文收集的测试集上的 RIMVFD 测试的 ROC 曲线

本文方法能够取得较好检测效果的原因在于使用了两段式弱分类器结构, 第一阶段所输出的实值分类结果被送入第二阶段进行更精确的分类, 使属于某个姿态区间的人脸目标可以更好地与姿态区间分割开来.

### 3.5 动态配置对系统性能的影响

本节分析动态配置策略对硬件实现代价、检测速度和精度的影响.

#### 1) 对子窗口扫描幅和缩放系数的动态配置

图 13, 14 所示为当改变子窗口扫描幅 ( $S_x$ ,  $S_y$ ) 和缩放系数  $r$  时检测速度和精度的变化情况. 可以看出, 随着  $S_x$ ,  $S_y$  和  $r$  的增大, 检测速度提高而检测精度下降, 原因是  $S_x$ ,  $S_y$  和  $r$  的增大将减少一幅

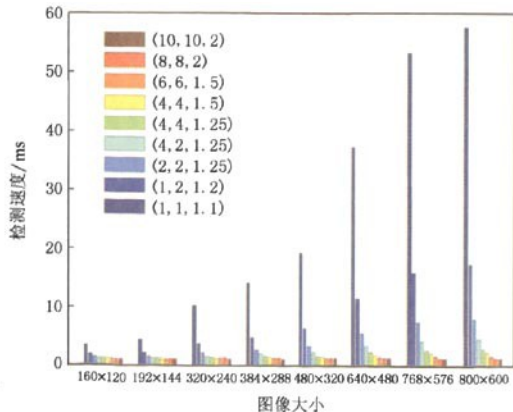


图 13 改变  $S_x$ ,  $S_y$  和  $r$  对检测速度的影响

图像中待检测子窗口的数量, 从而跳过一些可能存在于人脸目标的子窗口.

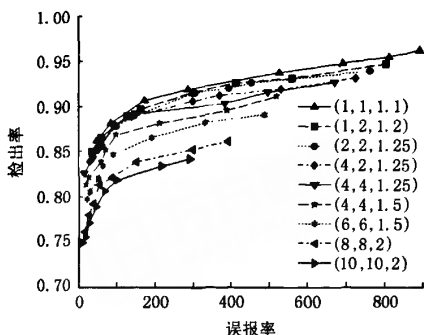


图 14 改变  $S_x$ ,  $S_y$  和  $r$  对检测精度的影响 (320×240 图像)

#### 2) 对检测节点、强分类器和弱分类器数量的动态配置

检测节点、每个节点中强分类器和每个强分类器中弱分类器的数量将影响到硬件实现代价, 图 15 所示为检测器结构所需的 ALUT 数量随弱分类器数量的变化情况. 可以看出, 二者之间是一种近似线性的关系, 当硬件资源足够多时, 可以配置较多的弱分类器来实现更高的检测精度和更细粒度的姿态区间分类.

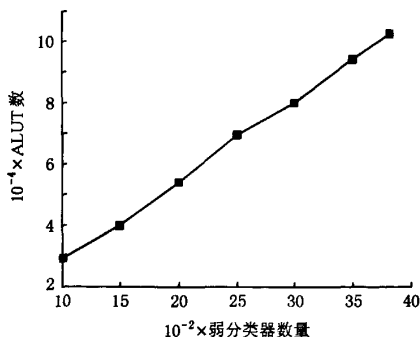


图 15 ALUT 数量随弱分类器数量的变化情况

## 4 结 论

针对  $[-90^\circ, 90^\circ]$  ROP 和  $360^\circ$  RIP 范围内的 RIMVFD 问题, 本文提出了一种精确分类的视角无关人脸检测方法和一种可重构硬件体系结构. 由多个检测节点组成的树形检测器框架结构实现了对人脸目标的由粗到细的 RIP 和 ROP 姿态识别与分类, 采用两段式 Boosting 方法提高了对人脸姿态分类的精确性. 实验结果表明, 本文方法与相关工作

相比具有较高的检测精度;另外,由于本文提出的RIMVFD硬件体系结构是高度并行和流水化的,与软件实现方案相比具有很高的执行时间加速比.本文的主要贡献在于:通过树形检测器框架结构和精确分类 Boosting 方法实现了 $[-90^\circ, 90^\circ]$ ROP和 $360^\circ$ RIP范围内的RIMVFD;通过充分挖掘算法的并行性,大大降低了检测过程的执行时间;通过对硬件体系结构模板进行动态配置,实现了硬件实现代价、检测精度和速度之间的折中与平衡.

### 参考文献(References):

- [1] Yang M H, Kriegman D J, Ahuja N. Detecting faces in images: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(1): 34-58
- [2] Rowley H A, Baluja S, Kanade T. Neural network-based face detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(1): 23-38
- [3] Schneiderman H, Kanade T. A statistical method for 3D object detection applied to faces and cars[C] //Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hilton Head, 2000: 746-751
- [4] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features [C] //Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, 2001: 511-518
- [5] Wu Bo, Huang Chang, Ai Haizhou, *et al.* A multi-view face detection based on real Adaboost algorithm [J]. Journal of Computer Research and Development, 2005, 42(9): 1612-1621 (in Chinese)  
(武 勃, 黄 畅, 艾海舟, 等. 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)
- [6] Huang C, Ai H Z, Li Y, *et al.* Vector boosting for rotation invariant multi-view face detection [C] //Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, 2005: 446-453
- [7] Schapire R E, Singer Y. Improved boosting algorithms using confidence-rated predictions [J]. Machine Learning, 1999, 37(3): 297-336
- [8] Mita T, Kaneko T, Hori O. Joint Haar-like features for face detection [C] //Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, 2005: 1619-1626
- [9] Baluja S, Sahami M, Rowley H A. Efficient face orientation discrimination [C] //Proceedings of International Conference on Image Processing, Singapore, 2004: 589-592
- [10] Dong Yazhuo, Dou Yong, Liu Mingzheng. A design space exploration algorithm for sliding-window operations [J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(5): 631-636 (in Chinese)  
(董亚卓, 窦 勇, 刘明政. 面向滑动窗口应用的设计空间探索方法[J]. 计算机辅助设计与图形学学报, 2008, 20(5): 631-636)
- [11] Li Baofeng, Dou Yong. Memory optimizing scheme and FPGA implementation of bit plane encoder [J]. Journal of Computer-Aided Design & Computer Graphics, 2008, 20(12): 1535-1540 (in Chinese)  
(李宝峰, 窦 勇. 位平面编码存储优化算法及 FPGA 设计[J]. 计算机辅助设计与图形学学报, 2008, 20(12): 1535-1540)
- [12] Theodoridis T, Link G, Vijaykrishnan N, *et al.* Embedded hardware face detection [C] //Proceedings of the 17th IEEE International Conference on VLSI Design, Las Vegas, 2004: 133-138
- [13] Yang M, Crenshaw J, Augustine B, *et al.* Face detection for automatic exposure control in handheld camera [C] // Proceedings of the 4th IEEE International Conference on Computer Vision Systems, New York, 2006: 17-24
- [14] Viola P, Jones M J. Robust real-time face detection [J]. International Journal of Computer Vision, 2004, 57(2): 137-154
- [15] Mitéran J, Matas J, Bourennane E, *et al.* Automatic hardware implementation tool for a discrete Adaboost-based decision algorithm [J]. EURASIP Journal on Applied Signal Processing, 2005, 2005(1): 1035-1046
- [16] Yu W, Xiong B, Chareonsak C. FPGA implementation of AdaBoost algorithm for detection of face biometrics [C] // Proceedings of IEEE International Workshop on Biomedical Circuits and Systems, Singapore, 2004: 17-20

作者: [徐金波](#), [窦勇](#), [Xu Jinbo](#), [Dou Yong](#)  
作者单位: [国防科学技术大学并行与分布处理国家重点实验室, 长沙, 410073](#)  
刊名: [计算机辅助设计与图形学学报](#) **ISTIC EI PKU**  
英文刊名: [JOURNAL OF COMPUTER-AIDED DESIGN & COMPUTER GRAPHICS](#)  
年, 卷(期): 2010, 22(1)  
被引用次数: 0次

## 参考文献(16条)

1. [Yang M H, Kriegman D J, Ahuja N](#) [Detecting faces in images: a survey](#) 2002(1)
2. [Rowley H A, Baluja S, Kanade T](#) [Neural network-based face detection](#) 1998(1)
3. [Schneiderman H, Kanade T](#) [A statistical method for 3D object detection applied to faces and cars](#) 2000
4. [Viola P, Jones M](#) [Rapid object detection using a boosted cascade of simple features](#) 2001
5. 武勃, 黄畅, 艾海舟, 劳世斌 [基于连续Adaboost算法的多视角人脸检测](#) [期刊论文] - [计算机研究与发展](#) 2005(9)
6. [Huang C, Ai H Z, Li Y](#) [Vector boosting for rotation invariant multi-view face detection](#) 2005
7. [Schapire R E, Singer Y](#) [Improved boosting algorithms using confidence-rated predictions](#) 1999(3)
8. [Mita T, Kaneko T, Hori O](#) [Joint Haar-like features for face detection](#) 2005
9. [Baluja S, Sahami M, Rowley H A](#) [Efficient face orientation discrimination](#) 2004
10. [董亚卓, 窦勇, 刘明政](#) [面向滑动窗口应用的设计空间探索方法](#) [期刊论文] - [计算机辅助设计与图形学学报](#) 2008(5)
11. [李宝峰, 窦勇](#) [位平面编码存储优化算法及FPGA设计](#) [期刊论文] - [计算机辅助设计与图形学学报](#) 2008(12)
12. [Theocharides T, Link G, Vijaykrishnan N](#) [Embedded hardware face detection](#) 2004
13. [Yang M, Crenshaw J, Augustine B](#) [Face detection for automatic exposure control in handheld camera](#) 2006
14. [Viola P, Jones M J](#) [Robust real-time face detection](#) 2004(2)
15. [Mitérán J, Matas J, Bourennane E](#) [Automatic hardware implementation tool for a discrete Adaboost-based decision algorithm](#) 2005(1)
16. [Yu W, Xiong B, Chareonsak C](#) [FPGA implementation of AdaBoost algorithm for detection of face biometrics](#) 2004

本文链接: [http://d.g.wanfangdata.com.cn/Periodical\\_jsjfszsjtxxxb201001024.aspx](http://d.g.wanfangdata.com.cn/Periodical_jsjfszsjtxxxb201001024.aspx)

授权使用: 陕西理工学院(sxlgxy), 授权号: 9708e212-7bbc-4c73-966f-9df201121e50

下载时间: 2010年9月15日