# OpenCV adaboost程序训练并检测目标

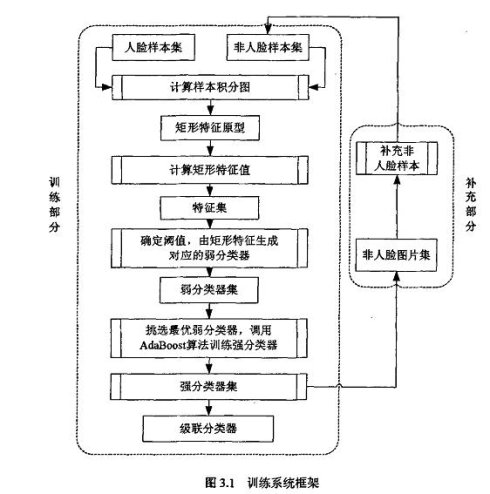
Adaboost的原理简述：每个Haar特征对应看一个弱分类器，但并不是任伺一个Haar特征都能较好的描述人脸灰度分布的某一特点，如何从大量的Haar特征中挑选出最优的Haar特征并制作成分类器用于人脸检测，这是AdaBoost算法训练过程所要解决的关键问题。

Paul Viola和Michael Jones于2001年将Adaboost算法应用于人脸检测中,其基本思想是针对不同的训练集训练同一个分类器（弱分类器），然后把这些不同训练集上的得到的分类器联合起来，构成一个最终的强分类器。Adaboost 算法中不同的训练集是通过调整每个样本对应的权重来实现的。开始时，每个样本对应的权重是相同的，对于h1 分类错误的样本，加大其对应的权重； 而对于分类正确的样本， 降低其权重， 这样分错的样本就被突出出来，从而得到一个新的样本分布 U2 。在新的样本分布下，再次对弱分类器进行训练，得到弱分类器 h2 。依次类推，经过 T 次循环，得到 T 个弱分类器，把这 T 个弱分类器按一定的权重叠加（boost）起来，得到最终想要的强分类器。

OpenCV Adaboost系统：

训练系统总体框架，由“ 训练部分”和 “ 补充部分”构成。依据系统框架，本文的训练系统可分为以下几个模块:

1. 以本集为输入，在给定的矩形特征原型下，计算并获得矩形特征集;
2. 以特征集为输入，根据给定的弱学习算法，确定闽值，将特征与弱分类器一一对应，获得弱分类器集;
3. 以弱分类器集为输入， 在训练检出率和误判率限制下， 使用A d a B o o s t 算法挑选最优的弱分类器构成强分类器;
4. 以强分类器集为输入，将其组合为级联分类器;
5. 以非人脸图片集为输入，组合强分类器为临时的级联分类器，筛选并补充  
   非人脸样本。

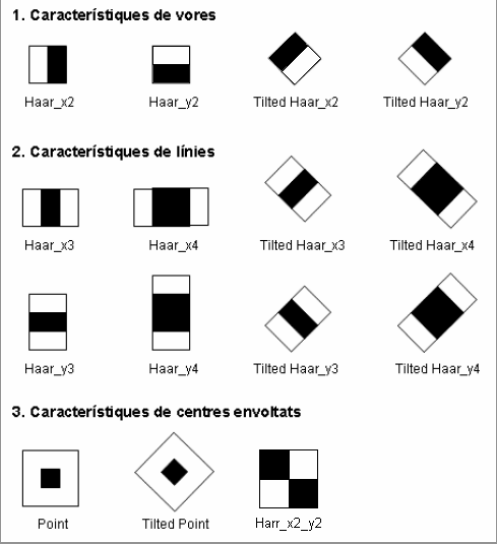


训练过程分为3个步骤：

首先需要提取Haar特征；然后将Haar特征转化成对应的弱分类器；最后从大量的弱分类器中迭代选择出最优弱分类器。

1. 提取Haar特征

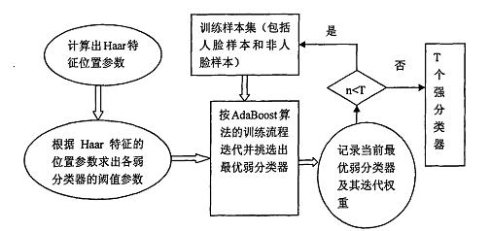
常用的Haar特征：



1. 生成弱分类器

每一个Haar特征都对应着一个弱分类器，每一个弱分类器都是根据它所对应的Haar特征的参数来定义的。利用上述Haar特征的位置信息，对训练样本进行统计就可以得到对应的特征参数。AdaBoost算法中所训练的弱分类器是任何分类器，包括决策树，神经网络，隐马尔科夫模型，如果弱分类器是线性神经网络，那么AdaBoost算法每次将构造多层感知器的一个节点。

1. AdaBoost算法训练过程就是挑选最优弱分类器，并赋予权重过程，



可能遇到的问题：

1.如果跑到某一个分类器时，几个小时也没有 反应，而且显示不出训练百分比，这是因为你的负样本数量太少，或者负样本的尺寸太小，所有的负样本在这个分类器都被reject了，程序进入不了下一个循环，果断放弃。解决方法：负样本尽量要大一些，比如正样本是40\*15，共300个,负样本是640\*480，共500个。

2.读取样本时报错：Negative or too large argument of CvAlloc function，网上说这个错误是因为opencv规定单幅iplimage的内存分配不能超过10000。把负样本的数量减少，尺寸加大，这个问题就解决了。

3.训练的过程可能经常出错，耐心下来不要着急，训练MRI分类器的时候失败了无数次。失败的时候有两件事可以做，第一，调整正负样本的数量，再试。第二，调整负样本的大小。

样本创建:

训练样本分为正例样本和反例样本，其中正例样本是指待检目标样本，反例样本指其它任意图片。

负样本: 样本可以来自于任意的图片，但这些图片不能包含目标特征

负样本由背景描述文件来描述。背景描述文件是一个文本文件，每一行包含了一个负样本图片的文件名（基于描述文件的相对路径）。该文件创建方法如下：

**for /f "delims=" %%i in ('dir /b/a/-d/s %neg\_path%')do (**

**if exist %%i\ (echo skip) else (echo %%i>>negtive\_info.dat))**

正样本: 对于正样本，通常的做法是先把所有正样本裁切好，并对尺寸做规整. 由于HaarTraining训练时输入的正样本是vec文件，所以需要使用OpenCV自带的CreateSamples程序.

1) 制作一个正样本描述文件，用于描述正样本文件名（包括绝对路径或相对路径），正样本数目以及各正样本在图片中的位置和大小。典型的正样本描述文件如下：

posdata/1(10).bmp 1 0 0 20 20　　　　  
posdata/1(11).bmp 1 0 0 20 20  
posdata/1(12).bmp 1 0 0 20 20

该文件创建方法如下(归一化为20X20)：

**for /f "delims=" %%i in ('dir /b/a/-d/s %pos\_path%')do (**

**if exist %%i\ (echo skip) else (echo %%i 1 0 0 20 20>>negtive\_info.dat))**

2) 运行CreateSamples程序。

createsamples -info posdata.dat -vec pos.vec -num 50 -w 20 -h 20

训练分类器:

haartraining.exe –data cascade -vec pos.vec -bg negdata.dat -npos 49 -nneg 49 -mem 200 -mode ALL -w 20 -h 20

利用训练好的分类器进行目标检测:

performance.exe -data cascade -info test.dat -w 20 -h 20 -rs 30

有用的脚本: 