

基于自举算法和支持向量机的人脸检测系统

胡 凯, 费耀平

HU Kai, FEI Yao-ping

中南大学 信息科学与工程学院, 长沙 410083

College of Information Science and Engineering, Central South University, Changsha 410083, China

E-mail: kevonhuo@163.com

HU Kai, FEI Yao-ping. Face detecting system based on AdaBoost algorithm and Support Vector Machine. Computer Engineering and Applications, 2008, 44(12): 199-203.

Abstract: For there were some tough problems in face detection under various background in gray level images, like too time-consuming and high false positive rate, this paper developed an applicable cascade face detecting system. First part of the system adopted extended Haar-like features combined with AdaBoost algorithm, which made performance of classifying better than using original Haar-like features. Second part of the system adopted vision logic, a strategy from general to detail to sample images, and introduced the conception of front-right face image pixel confidence, and took Support Vector Machine as learning algorithm. This part ensured the good accuracy of the detection. This system performed well on real image test. This was an applicable face detecting system.

Key words: face detection; Haar-like feature; AdaBoost; Support Vector Machine(SVM); pixel confidence

摘 要: 针对复杂背景下的灰度图像人脸检测存在计算量大且负检率高等问题, 提出了一种有较好可用性的层级递进的人脸检测系统。系统第一部分采用扩展的 Haar 型特征并结合自举算法, 使其分类性能要优于原始的 Haar 型特征。在系统的第二部分, 采用从粗到细的视觉处理逻辑对图像采样, 并提出了正面直立人脸的像素值的置信度的概念, 且以支持向量机作为学习算法, 使系统具有良好的检测性能。该系统在实际应用图像的测试中取得良好效果, 具有可用性。

关键词: 人脸检测; Haar 型特征; 自举算法; 支持向量机; 像素置信度

文章编号: 1002-8331(2008)12-0199-05 **文献标识码:** A **中图分类号:** TP751.1

1 引言

人的生物特征如指纹、虹膜、人脸等具有不可复制、携带方便、鉴别性强等特征, 是理想的辨别依据。其中人脸识别由于其非接触性、非侵犯性等特点是最自然直接的身份辨别手段。而人脸检测是人脸识别的第一步。其算法的性能直接影响到整个人脸识别系统的识别准确率和性能。

本文研究的是在有复杂背景的灰度图像中检测人脸。由于光照、姿态、大小、表情等因素造成人脸的模式变得复杂, 人脸检测就变成一个具有挑战性的任务。这些人脸模式的变化造成系统需对图像作全面的扫描, 这样的扫描使人脸检测系统运算量巨大。要在可接受的时间内完成各种各样的图像精确的人脸检测任务变得十分困难。

人脸检测发展至今也有比较成功的方法。如文[1, 2]描写的算法有很高的精确度, 适合那些对时间性能要求不高的应用。文[3]对人脸模式在特征脸子空间中的分布进行建模来完成人脸的检测。文[4]将神经网络应用于人脸检测。Viola 等^[5]用过完备 Haar 型特征建立一个层级自举的人脸检测系统, 它用自举学习算法(AdaBoost)设计层级结构, 其检测速度较快但精确度不如前面所提及的方法。本文兼顾时间效率和精确度的要求,

将运算简单的 Haar 型特征进行扩展, 并结合了支持向量机(SVM)神经网络, 设计了一个由两部分组成的层级分类的人脸检测系统。

本文的创新点:

(1) 扩展了 Haar 型特征, 并采用 AdaBoost 算法对特征进行选择排序。

(2) 计算正面直立人脸的像素值的置信度, 它反映了像素值的稳定性。

(3) 采用从粗到细的视觉处理逻辑对图像采样, 并用经过训练的 SVM 对图像分类。

2 系统概览

出于分类能力和计算量的考虑系统分为两部分。第一部分可以快速地排除非人脸区域的输入, 第二部分计算量较第一部分有所增加, 主要用于较精确地将那些第一部分没有能力去除的非人脸输入排除。

第一部分采用扩展的 Haar 型特征, 并用 AdaBoost 算法对特征进行选择排序, 再取分类能力较强的特征进行分层。本文将分为三层, 结构为 3-30-60。第二部分, 首先要从训练数据

作者简介: 胡凯(1982-), 男, 在读硕士, 主要研究领域为图像处理, 模式识别; 费耀平(1959-), 男, 教授, 主要研究方向为图形图像处理、模式识别、数据挖掘。

收稿日期: 2007-08-08

修回日期: 2007-11-16

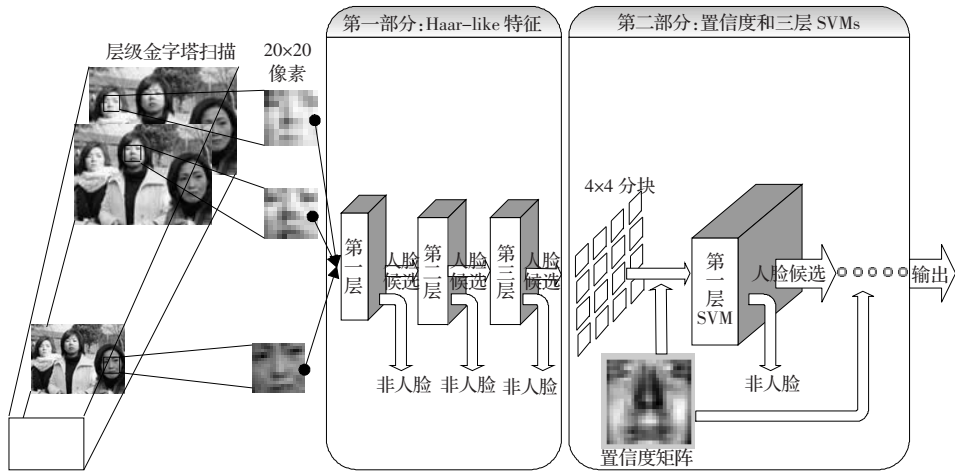


图1 系统逻辑结构

中计算人脸图像的置信度权重用于后面的 SVM 特征输入。SVM 的输入采用人眼视觉的处理逻辑,先将 24x24 图像作分割,本文分别做了 4x4、8x8、12x12 的三层分割,将每一块的像素置信值 C_n 组成向量作为输入。这一部分的分类器同样要分层,即从粗到细本文分为三层 4x4、8x8、12x12。系统结构如图1。

本系统也如其他类似系统采用金字塔扫描算法。以固定窗口对图像进行扫描,再以变化因子变化图像大小后以同样的方法进行扫描。本文采用的变化因子为 1.18。变化因子越大,则扫描越粗,耗费时间越少,但漏检率也越高。

3 第一部分:扩展的 Haar 型特征

为了减少类内元素的差别而增大类间元素的差别,常常以特征值而不是原始像素值作为学习算法的输入。这样会简化分类的算法且增加分类的正确率。但是从有限的原始像素值计算出的特征值往往是复杂的。所以一种既有效又简单的特征计算方法变得十分重要。Papageorgiou 等^[6,7]提出过完备 Haar 型特征集。且 Viola 等在文[5]中提出了一种该特征值的快速计算方法。本文在其特征集上进行扩展,增加了三种特征。

3.1 特征族

在 Viola 和 Jones^[5]的人脸检测算法中涉及了四种 Haar 型特征(如图 2 中 A、B、C、D 所示)。在图 2 中,特征 A 和特征 B 为边缘特征,它由水平或竖直对齐且相邻的两个长宽相等的矩形区域组成,它反映的是两块相邻的矩形区域像素值总和的差。特征 C 为线特征,它由竖直对齐且相邻的 3 个长宽相等矩形区域组成,它反映的是两块外部的矩形区域和中间的矩形区域像素值总和的差值。最后一个特征 D 是由 4 个相邻的长宽相等的矩形组成,它反映的是对角区域的像素和差值。

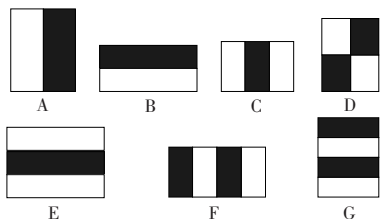


图2 扩展的 Haar 型特征

本文在此基础上又增加了三种特征(如图 2 中 E、F、G 所

示)。特征 E 为线特征,与特征 C 相似,但体现的是水平边界特征。特征 F 为栅栏特征,由水平对齐且相邻的 4 个长宽相等矩形区域组成的竖直方向的栅栏状特征,它反映的是相互间隔的矩形区域像素值总和的差。特征 G 与特征 F 相似,但体现的是水平方向的栅栏特征。这些特征定义如下:

$$feature_n = F(style_n, sx, sy, h_n, w)$$

其中 $style$ 是特征类型(A、B、C、D、E、F、G); sx 为整个特征矩形区域左上角 x 坐标; sy 为整个特征矩形区域左上角 y 坐标; h 为整个特征矩形区域的高; w 为整个特征矩形区域的宽,高和宽的单位都是像素。

本文采用的是大小为 24x24 的检测窗口。在 24x24 的窗口中整个特征集十分庞大,共有 134 736 种特征,远远超过整个窗口像素个数 576。所以这是一个过完备特征集。完备特征集对图像表达更为丰富,从而提高了分类的正确率。文[8]也对 Viola 等^[5]的特征进行了扩展,因为它对特征加入了长宽比的限制,在 24x24 的窗口中特征数有 117 941 种。

3.2 特征的计算

本文对于特征的计算借助积分图像来完成。积分图像能使矩形区域的像素求和十分快捷。积分图像是对原图像的求和。在积分图像(x, y)位置的值是所有在左上方原图像像素值的总和:

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

$ii(x, y)$ 是积分图像像素值, $I(x', y')$ 是原图像像素值。采用公式(1)和(2)则只遍历一遍原始图像即可得到积分图像。

$$h(x, y) = h(x, y-1) + I(x, y) \tag{1}$$

$$ii(x, y) = ii(x-1, y) + h(x, y) \tag{2}$$

其中 $h(x, y)$ 为原图像第 x 列的从上至下 y 个像素的和。

借助积分图像,任何一个矩形区域的像素值总和都可以由积分图像中的 4 个位置的值作加减运算得到:

$$rec(x, y, h, w) = ii(x+w-1, y+h-1) - ii(x-1, y+h-1) - ii(x+w-1, y-1) + ii(x-1, y-1)$$

其中参数 x, y 为矩形区域左上角坐标, h 为矩形的高, w 为矩形的宽。可见无论特征的矩形区域大小,边缘特征只需要 8 次加减运算,线特征需 9 次加减运算,栅栏特征需 11 次加减运算。这种特征的计算方式与其他方法相比具有明显的速度优势。

3.3 特征的选择和排序

每一种特征都可以作为学习算法的输入,在训练数据库上

加以训练从而产生分类器。一种特征的值在人脸集和非人脸集中的分布近似高斯分布。用一种特征能构造一种简单的弱分类器。对于每一种特征的分类器,本文选取的是能使在训练数据集上正检率达到最大的阈值,而负检率可通过后面的层级降低(见本文 3.5 节)。由一种特征构成的分类器描述如下:

$$h_n(x) = \begin{cases} 1 & p_n \cdot \text{feature}_n(x) < p_n \cdot \theta_n \\ 0 & \text{otherwise} \end{cases}$$

其中 p_n 指示不等式方向, $\text{feature}_n(x)$ 为 feature_n 在 24×24 像素的图像 x 上的值, θ_n 为阈值。

因此,本文将产生 134 736 个弱分类器。对于这些基于某一个特征的弱分类器的选择已经存在许多算法,算法的讨论见文[5]。本文将采用自举(AdaBoost)方法对其进行排序选择,从而产生分类速度和效果都比较理想的强分类器。自举学习算法将弱分类器的分类结果通过权重结合起来以提高弱分类器的分类能力降低分类错误率。在自举算法中,为了提高弱分类器的分类能力,在完成第一轮学习后,所有的训练样本的权重都需重新计算,从而强调被上一个分类器错误分类的样本,使学习越来越难。因此自举算法能有效地从大量的分类特征中找到那一小部分最好的分类特征。且 Freund 和 Schapire 在文[9]中已经证明了自举学习算法的泛化能力。自举算法如下:

(1) 训练样本 $(x_1, y_1), \dots, (x_n, y_n)$, 其中 $y_i=0, 1, 0$ 和 1 分别表示非人脸和人脸。

(2) 初始化权重 $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$, m 为非人脸样本个数, n 为 人脸样本个数。

(3) For $t=1, \dots, T$:

1. 归一化权重值

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

2. 对于每一个 feature_j , 训练只有一个特征的弱分类器 h_j , 并误率如下:

$$e_j = \sum_i w_i |h_j(x_i) - y_i|$$

3. 对于 e_j 进行排序, 取最小值记为 e_t , 对应分类器记为 h_t 。

4. 更新权重值:

$$w_{t+1,i} = w_{t,i} \beta_t^{1-r_i}$$

其中 $r_i = \begin{cases} 0 & x_i \text{ is classified correctly} \\ 1 & \text{otherwise} \end{cases}$

$$\beta_t = \frac{e_t}{1-e_t}$$

(4) 产生强分类器:

$$h(x) = \begin{cases} 1 & \sum_{i=1}^T \alpha_i h_i(x) \geq \frac{1}{2} \sum_{i=1}^T \alpha_i \\ 0 & \text{otherwise} \end{cases}$$

其中 $\alpha_i = \log \frac{1}{\beta_i}$ 。

3.4 AdaBoost 的学习结果

对于用自举学习算法产生的强分类器显然不能包括全部的弱分类器。但包括的弱分类器越多也就是 T 越大, 则该强分类器的分类的正检率越高, 负检率越低。本文用相同的样本数据, 对弱分类器数量和正检率及负检率的关系作了对比实验,

实验结果见本文 5.2 节。

在 Viola[®] 的人脸检测算法中涉及了四种 Haar 型特征。本文在此基础上增加了三种特征。在自举算法选择出来的特征中特征 E (如图 2 所示) 位于第一位。本文对 Viola[®] 中的四种特征和本文采用的七种特征的分类能力在同一批样本数据中作了比较。本文采用的七种特征的分类能力明显高于仅有四种特征。七种特征扩大了原有的四类特征的特征空间, 对图像的表达更为丰富, 从而分类能力就会有相对的提高。实验结果见 5.2 节。

3.5 分类器的层级结构

在人脸检测中大于 99% 的子窗口输入都为非人脸。所以如何快速排除这些非人脸输入成了人脸检测中提高效率的关键问题。在这些非人脸输入中并不是所有的输入都要经复杂的计算才能作出判断, 有些只需要经过简单的计算就能排除。因此就引出了分层过滤的方法。将那些很容易判断的非人脸的输入用简单的计算排除, 剩下的比较难以判断的输入才采用较复杂的算法。这种方式明显提高了计算效率, 减少了计算时间。本文就采用这种层级的方法。整个层级分类器的负检率计算如下:

$$F = \prod_{i=1}^K f_i$$

其中 f_i 是第 i 层的负检率, K 为特征数。

显然以层级的方式负检率下降非常快。在第一层中用自举算法选择出来的前几个特征组成的强分类器对输入进行过滤。这种过滤要求计算量少且去除了大部分非人脸输入。因为采用的特征数增多虽然负检率降低但计算量增加, 3 个特征能满足第一层的计算量和负检率要求, 所以本文第一层采用 3 个特征。出于负检率的要求第二层本文采用 30 个特征, 第三层 50 个特征。本文在同一批非人脸样本上实验得到负检率: 第一层 4.00%, 两层 0.65%, 三层一起为 0.40%。显然通过这三层后 99.60% 的非人脸输入都被排除。剩下的部分本文将用比较复杂的计算进行分类。

4 第二部分: 支持向量机

神经网络能帮助解决那些不能精确建立数学模型的问题, 以它为工具做目标检测已经有很长的历史了(见文[4])。神经网络发展至今已十分成熟, 且产生了许多分支, 支持向量机(SVM)就是其中一种常被目标检测算法所采用的神经网络。支持向量机在结构风险最小化方面独辟蹊径, 在泛化性能上使其他算法望尘莫及。

4.1 置信度的计算

Henry A. Rowley 在文[4]中, 将 20×20 的图像按照一定比例分割, 分割后每一块的像素值总和组成向量作为神经网络的输入。但人脸是一个弹性物体, 随着表情的差异以及个体之间的面貌差异, 图像的像素值差别较大。因为人脸的模式还是有相对固定的部位, 某些位置的像素值变化相对稳定。所以本文提出了像素置信度的概念。利用置信度计算图像各分块的平均置信值 C_i 。在本文分别对不做置信度处理和做处理进行了对比实验。实验结果见 5.2 节。实验表明加入了置信度概念后, SVM 的分类能力得到了明显的提高。本文将对 24×24 人脸图像中各位置的像素值置信度进行计算。计算像素值置信度如下:

(1) 计算平均人脸

$$a(x, y) = \frac{1}{T} \sum_{i=1}^T I_i(x, y) - \mu_i$$

$$\mu_t = \frac{1}{h \cdot w} \sum_{x \leq w, y \leq h} I_t(x, y)$$

其中 T 为训练样本数, $I_t(x, y)$ 为样本 t 在 (x, y) 位置的像素值, μ_t 为样本 t 的平均值, h 为样本图像高(本文为 24), w 为样本图像的宽(本文为 24)。本文减去 μ_t 是为了能快速去除不同光照对图像的影响。

(2)对每一个像素计算其置信度

$$W(x, y) = \frac{T}{\sum_{t=1}^T |I_t(x, y) - \mu_t - a(x, y)|}$$

因此,图像中像素稳定性较高的占有较大权重。图 3 为人脸的置信度图,越亮的部分置信度越大。在对神经网络进行输入时,如 Henry A. Rowley 先把图像划分成面积相等的数块^[4],然后计算每一块的置信值 C_i (见式(3)),将 C_i 组成向量 $[C_1, C_2, \dots, C_n]$ 作为支持向量机的输入。



图3 人脸置信度谱

$$C_i = \frac{\sum_{x'_i \leq x \leq x''_i, y'_i \leq y \leq y''_i} [I(x, y) - \mu] \cdot W(x, y)}{\sum_{x'_i \leq x \leq x''_i, y'_i \leq y \leq y''_i} W(x, y)} \quad (3)$$

其中 x'_i, x''_i 为这第 i 块的 x 的起止坐标, y'_i, y''_i 为这第 i 块的 y 的起止坐标。若不引入置信度,只采用块的像素平均值,则计算如下:

$$C'_i = \frac{\sum_{x'_i \leq x \leq x''_i, y'_i \leq y \leq y''_i} [I(x, y) - \mu]}{(y''_i - y'_i) \cdot (x''_i - x'_i)}$$

4.2 图像的划分

现在还有一个问题就是输入图像应被怎么样划分成块。按照人眼视觉的分析图像的过程也是这样,在辨认物体的时候先粗看发现候选目标后再细看进行进一步辨识。本文也引用了这样的处理逻辑。划分得越细则分类会越精确,但计算量会越大,划分得越粗则分类的负检率就会越高,但运算开销会越小。综合这两种情况本文采取层级递进的方法。第一层将图像作 4×4 的划分,即以 16 维置信值向量作为支持向量机输入,用正负训练样本集 T, F_1 对其进行训练,以惩罚因子 $C=500$ 训练得到 46 个支持向量。对该支持向量机在非人脸样本上作测试,保留误检样本作为下一层的训练负样本 F_2 。第二层将图像作 8×8 划分得到 64 维置信值向量,对其以惩罚因子 $C=600$ 训练得到 82 个支持向量。按第二层的测试方法产生第三层的负训练样本。第三层将图像作 12×12 划分得到 144 维置信值向量,对其训练,以惩罚因子 $C=600$ 训练得到 153 个支持向量。显然划分得越细即分类越细致,则输入的维数越大支持向量会越多,那么计算量就会大大增加。但并不是所有的输入图像都需要经过比较细的划分才能正确分类,因此如此层级递进将会大大缩减计算开销。

5 实验结果和讨论

5.1 实验准备

本文所采用的人脸测试和训练数据是从许多人脸数据库及网络中剪切得来。这些数据库包括 BioID、VALID Database、ORL 等。剪切方法是先将正面直立人脸从眼眶上沿开始,左右各到两眼的外眼角,下沿到嘴附近,使整个剪切成正方形。这种取样本的方法尽可能排除了其他脸部不稳定的部分,如发型、头发的遮挡、衣领的高低和颜色、下巴的形状和胡须等。最后,把剪切成的灰度图像映射成 24×24 (如图 4 所示)。非人脸图像是来自网络的不包括人脸的风景图片,对其用 24×24 的子窗口进行扫描从而得到非人脸样本。



图4 人脸训练数据

5.2 实验内容

本文对用 AdaBoost 产生的强分类器(包括本文所提出的三种特征)的阈值进行调整从而产生了 ROC 曲线(如图 5 所示),该图反映了弱分类器数量和正检率、负检率的关系。

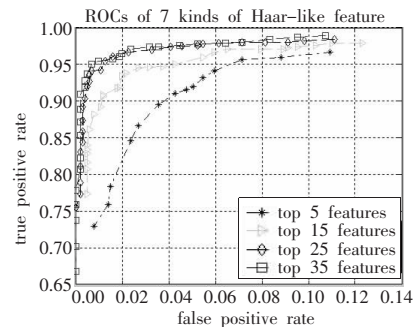


图5 对于不同特征数量的 7 类 Haar 型特征的 ROC 曲线

从图 5 可以看出特征数越多分类器性能越好,但分类性能的差别随着特征数增多而减小,所以必须平衡好分类性能和运算时间。本文采用 3-30-60 的层次结构正是出于这样的考虑。

本文对 Viola 等采用的四种特征的分类器和加入本文的三种特征的分类器在同一批训练数据上的 ROC 曲线进行了对比(如图 6-8 所示)。

从图 6-8 可以看出本文的七种特征的分类器的分类性能要优于 Viola 等采用的四种特征的分类器。只有快速有效地把非人脸排除才能使人脸检测系统具有可用性。利用 Haar 型特征的计算优点,又对原来的性能有所提高正是为了满足系统的要求。

本文对采用置信度后的分类效果和不采用的分类效果做了对比实验。实验是在同一批训练和测试数据上进行的,且对图像作 8×8 的分割。图 9 表明采用置信度后的分类性能得到了提高。

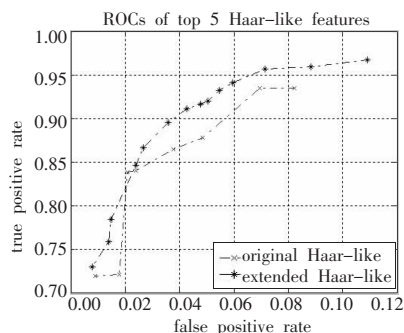


图6 前5位扩展和非扩展的 Haar 型特征 ROC 的对比

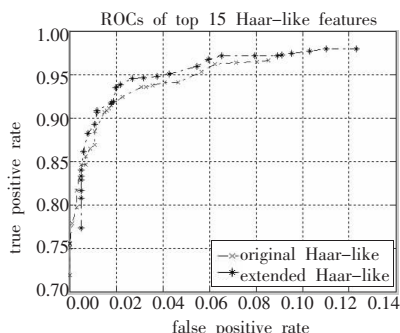


图7 前15位扩展和非扩展的 Haar 型特征 ROC 的对比

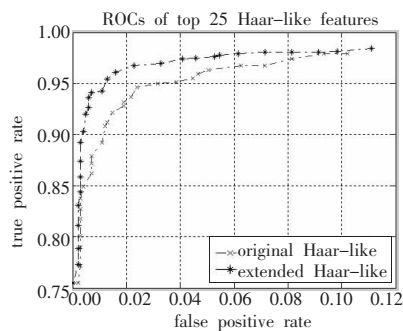


图8 前25位扩展和非扩展的 Haar 型特征 ROC 的对比

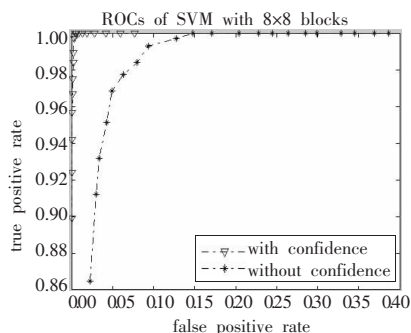


图9 采用和不采用置信度的 SVM 的 ROC

6 结束语

本文构建了一个层级递进的人脸检测系统,它由两部分组成:第一部分采用了易于计算扩展的 Haar 型特征以及自举学习算法以保证系统的时间要求。第二部分采用了以置信值为特征输入的 SVM 神经网络来保证系统对精度的要求。这每个部分都为三层,以减少计算开销。经对比实验可见扩展后的 Haar 型特征的分类性能要优于原始的 Haar 型特征,且采用由粗到细的对图像分割并结合像素置信度使系统分类精度大大提高。从来自 BioID、VALID Database、ORL 等人脸数据库和网络的包括正面直立人脸图像的检测结果来看,本系统具有较好的可用性。

(上接 174 页)

次对比仿真实验,分别说明 FDDD 和 FDEI 的优越性。

综上所述,本文方法结构新颖,抗噪性和实时性好,适用范围广且易于实现,具有重要的理论意义和潜在实用价值。

参考文献:

- [1] 邓振森,刘渝.正弦波频率估计的牛顿迭代方法初始值研究[J].电子学报,2007,35(1):104-107.
- [2] Aboutanios E,Mulgrew B.Iterative frequency estimation by interpolation on Fourier coefficients[J].Signal Processing,2005,53(4):1237-1242.
- [3] 齐国清.几种基于 FFT 的频率估计方法精度分析[J].振动工程学报,2006,19(1):86-92.
- [4] Gunawan D D S.Sinusoidal frequency estimation based on the

参考文献:

- [1] Roth D,Yang M,Ahuja N.A snowbased face detection [J].Neural Information Processing,2000,12.
- [2] Schneiderman H,Kanade T.A statistical method for 3D object detection applied to faces and cars[C]//Proc IEEE Computer Soc Conf on Computer Vision and Pattern Recognition,2000.
- [3] Mophaddom B,Pentland A D.Probabilistic visual learning for object representation[J].IEEE Trans PAMI,1997,19(7):696-710.
- [4] Rowly H A,Baluja S,Kanade T.Neural network-based face detection[J].IEEE Trans PAMI,1998,20(1):23-38.
- [5] Viola P,Jones M J.Rapid object detection using a boosted cascade of simple features[C]//Proc of International Conference on Computer Vision and Pattern Recognition,2001,1:511-518.
- [6] Mohan A,Papageorgiou C,Poggio T.Example-based object detection in images by components[J].IEEE Transactions on Pattern Analysis and Machine Intelligence,2001,23(4):349-361.
- [7] Papageorgiou C,Oren M,Poggio T.A general framework for object detection[C]//International Conference on Computer Vision,1998.
- [8] Lienhart R,Maydt J.An extended set of Haar-like features for rapid object detection[C]//Proc of the IEEE International Conference on Image Processing,2002,1:900-903.
- [9] Schapire R E,Freund V,Bartlett P.Boosting the margin:a new explanation for the effectiveness of voting methods[C]//Proceedings of the Fourteenth International Conference on Machine Learning,1997.

time derivative of the STFT phase response[J].Information,Communications and Signal Processing,2005:1452-1456.

- [5] 李靖,王树勋,汪飞.基于抽取技术的二维密集频率估计方法[J].电子学报,2005,33(9):1670-1674.
- [6] 郭敏,马远良,朱霆.基于数据融合及小波变换的医学超声图像去噪方法[J].计算机工程与应用,2006,42(6):13-15.
- [7] Stefan Franz,Sanjit K,Mitra G D.Frequency estimation using warped discrete Fourier transform[J].Signal Processing,2003,83(8):1661-1671.
- [8] 刘良兵,涂亚庆.基于多段分频等长信号融合的频率估计方法[J].计算机工程与应用,2008,44(11):170-175.
- [9] 杨露菁,余华.多源信息融合理论与应用[M].北京:北京邮电大学出版社,2006:132-141.
- [10] 胡广书.数字信号处理理论、算法与实现[M].2版.北京:清华大学出版社,2003:97-102.