

基于多级梯度能量描述的压缩域人脸检测

李晓华 沈兰荪

(北京工业大学信号与信息处理研究室 北京 100022)

摘要: 针对 JPEG2000 彩色图像, 提出了一种结合肤色和纹理信息, 直接在小波压缩域操作的人脸检测方法。该方法有 3 大特点: 首先, 提出了小波域人脸模式的多级梯度能量描述, 在有效表征脸部特点的同时, 可避免复杂的压缩域图像缩放操作, 首次较好地解决了压缩域人脸检测中尺寸未知的难点; 其次, 优化 YCbCr 彩色空间肤色模型, 提高肤色分割准确度; 最后, 在检测器的设计中, 将基于规则的简单分类器和基于神经网络的复杂分类器有机结合, 进一步加快处理速度。实验结果表明, 提出的方法是有效而快速的。

关键词: 人脸检测, JPEG2000, 肤色模型, 小波压缩域, 多级梯度能量

中图分类号: TP391.4 文献标识码: A 文章编号: 1009-5896(2005)12-1909-07

Face Detection in Compressed Domain Based on Multi-level Gradient Energy Presentation

Li Xiao-hua Shen Lan-sun

(Signal & Information Processing Lab, Beijing University of Technology, Beijing 100022, China)

Abstract In this paper, a fast face detection algorithm for JPEG2000 color images is presented, which combines both color and texture information in order to find a good balance between speed and detection reliability. The algorithm is designed to work directly on the wavelet compressed domain which possesses the following characters: First of all, the multi-level gradient energy presentation of face pattern is proposed, which not only can highlight the facial parts in possible face patterns, but also can address effectively the problem of unknown size in face detection in compressed domain and therefore avoid the complex resolution transform in arbitrary ratios; secondly, the skin-color model in YCbCr space is ameliorated to improve the reliability of skin segmentation; finally, a hierarchical detector which integrates the simply rule-based classifiers and complex neural network based classifier is designed to further improve the processing speed. Experimental results show that the proposed scheme is efficient and effective.

Key words Face detection, JPEG2000, Skin color model, Wavelet compressed domain, Multi-level gradient energy

1 引言

人脸检测作为人脸信息处理的重要内容之一, 在图像检索、视频场景分类、智能人机接口等诸多领域中有着极为广泛的应用价值, 已成为计算机视觉与模式识别领域的研究热点之一^[1]。人脸检测主要有两大类研究方法: 一类是基于肤色的方法^[2,3], 根据肤色和启发式信息定位人脸。优点是速度快, 对人脸姿态、表情、遮挡物(如眼镜, 耳环)等不敏感, 但误检率高且定位不精确。另一类是仅利用灰度信息的基于纹理的方法^[4,5], 把人脸检测看作 2 类模式识别问题, 即先对候

选子窗提取描述人脸模式的特征量, 后用模式识别方法(如神经网络、支撑向量机、模板匹配等)分类该特征矢量为人脸模式或非人脸模式。通常基于纹理的方法有较强的鲁棒性, 定位准确, 但计算复杂度高, 速度慢。显然, 针对彩色图像, 结合上述两种思路可获得较好的鲁棒性和较高的速度。

传统的人脸检测方法大多针对原始像素域设计, 应用于压缩格式图像时须先解压再检测, 计算量大、内存占用量多。为省去解压过程, 提高处理速度, 研究压缩域人脸检测很有实用价值。然而, 由于压缩域的固有限制, 研究难度较大。

2004-05-21 收到, 2005-04-13 改回

国家自然科学基金(60402036)、北京市自然科学基金(4042008)、教育部博士点科研基金(20040005015)和北工大博士科研启动基金资助课题

如针对JPEG和MPEG的I帧图像, Wang等^[6]提出利用DC系数中的肤色信息检测人脸,其特点是速度快,但具有与基于肤色检测法相似的缺点。文献[7]和文献[8]则试图将典型的基于纹理的方法从像素域映射到DCT域,但未能很好解决压缩域多尺度人脸检测问题。小波变换因其优良性能近年来受到广泛研究,已被最新国际压缩标准JPEG2000全面采用,可取得比JPEG标准更优良的压缩性能,因此研究基于小波压缩域的人脸检测将具有重要的意义和广泛的应用前景,但这方面研究尚未见之于文献。

本文针对JPEG2000彩色图像,提出了一种结合肤色和纹理信息、直接在小波压缩域操作的快速人脸检测方法。该方法的特点是:提出小波域人脸模式的多级梯度能量描述,有效解决了压缩域人脸检测中尺寸未知的问题;优化YCbCr彩色空间的肤色模型,使肤色分割更精确;最后,设计了一种集成基于规则的简单分类器和基于神经网络的复杂分类器的级联检测器,在保证检测率较高的同时,进一步加快处理速度。

文中提及小波压缩域或小波压缩数据时,均指基于小波变换的压缩格式图像(如JPEG2000)经部分解码(熵解码和反量化)后得到的多尺度塔式小波数据。其中LL是低频近似图像, LH_n, HL_n, HH_n 称为高频方向细节图像, n 越大,尺度越粗。在基于小波的图像压缩中,3级分解常常可以达到计算量和压缩效果的最佳折中。所以不失一般性,本文研究基于3级小波分解数据的人脸检测。

2 算法难点及整体思路

如何实现多尺度人脸检测,加快处理速度,是压缩域人脸检测急需解决的难点之一。在像素域,人脸尺寸未知的问题常通过固定人脸模板大小,按一定比率逐级缩小图像来解决。而在压缩域,要实现任意因子的图像缩放往往会引入较高的计算量,现有的几种压缩域人脸检测方法虽然已认识到该问题,但并未给出很好的解决方案。如Chua等^[7]设计了6个固定尺寸的人脸模板来避免DCT域图像缩放,但能检测的人脸尺寸范围有限($40 \times 40 \sim 80 \times 80$)。Luo等^[8]则采用一种可在DCT域执行任意因子缩放操作的有效算法,但相对于像素域缩放,该算法计算复杂度仍较高。本文在深入分析人脸模式在小波压缩域中表现特点的基础上,提出人脸模式的多级梯度能量描述,不但能有效表征各种尺寸人脸模式的固有特征,而且可避免复杂的压缩域图像缩放操作,首次较好地解决了压缩域多尺度人脸检测问题。

算法以JPEG2000图像部分解码后得到的DWT(Discrete Wavelet Transform)系数做为输入,按照肤色分割-人脸搜索的模式进行人脸检测,以矩形框形式输出检测结果的确切位置和尺寸。检测过程仅利用部分解压后的小波系数,省去小波

反变换,处理速度较高。算法总体框图如图1所示,由两个步骤构成:肤色区域分割和基于多级梯度能量描述的人脸检测。步骤1利用LL子带的(Y,Cb,Cr)值和肤色统计模型分割肤色区域。为尽可能精确地分割肤色区,给出了结合亮度分量的双肤色模型。采用中值滤波器滤除二值分割图像中的噪音点并填补空穴,用连通区域分析法对每个候选肤色区标号并计算其位置参数。步骤2即对候选肤色区,从Y分量的多尺度高频细节子带(LH_n, HL_n, HH_n)抽取多级梯度能量描述,用级联检测器在多尺度多位置搜索,精确定位人脸。

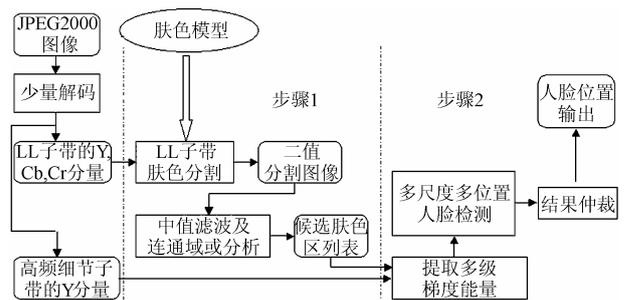


图1 算法整体框架

3 肤色区域分割

3.1 肤色模型

肤色模型的建立需选择彩色空间。不同彩色空间肤色的聚集程度(即紧致性)不同。本文选用YCbCr彩色空间,原因有二:(1)与其他彩色空间(RGB, HSV等)相比,肤色在该空间中分布最为集中^[3];(2), YCbCr空间被广泛应用于图像压缩中以消除空间相关性。对基于小波的压缩格式图像,彩色信息可在部分解压的LL子带中直接获取,无需任何中间变换。

为建立肤色模型,从包含不同人种和光照条件的90幅图像中人工裁剪其中的肤色区域,获得1,937,664点肤色样本。图2(a)为肤色样本在Cb-Cr平面的分布图,图中的黑色区对应于分布概率大于0.001的点。传统的肤色模型有两种:对Cb, Cr分离考虑的矩形肤色模型^[2](如图2(a)中的虚框, R_{Cb}, R_{Cr} 分别表示肤色在Cb, Cr分量的分布区间)和联合考虑Cb-Cr的适形肤色模型^[6](如图2(a)中的黑色区,用 $R_{Cb \& Cr}$ 表示)。矩形模型简单,直观,但包含了许多非肤色区,误分严重。适形模型是(Cb, Cr)的2D联合分布,分割效果较好。然而由于该模型未考虑亮度信息,在肤色分割时常会误把颜色较深的非肤色点分为肤色点。观察肤色点在YCbCr空间的分布,发现高亮度样本在Cb-Cr平面的分布覆盖区域较大,而低亮度对应于较小的覆盖区域。图2(b), 2(c)给出了 $Y \geq 128$ 和 $Y < 128$ 的肤色点在Cb-Cr平面的分布。可以看出,图2(b)与图2(a)的分布区基本一致,而图2(c)的分布区远远小于图2(a)。本文在适形肤色模型的基础上,适当引入亮度信息,

对高亮度和低亮度采用不同的肤色模型, 即 $R_{Cb\&Cr}^1$ ($Y \geq 128$) 和 $R_{Cb\&Cr}^2$ ($Y < 128$), 称为双适形模型。

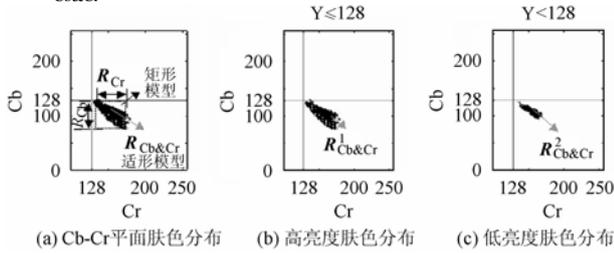


图 2 肤色模型

3.2 肤色分割

由于小波压缩域只有LL子带具有彩色信息, 因此肤色分割是基于LL子带进行的。对于LL子带每一点 (x,y) , 若 $Y(x,y) \geq 128$ && $[Cb(x,y), Cr(x,y)] \in R_{Cb\&Cr}^1$ 或 $Y(x,y) < 128$ && $[Cb(x,y), Cr(x,y)] \in R_{Cb\&Cr}^2$, 则分类为肤色点, 否则为非肤色点。实验中对 56 幅不同人种不同光照条件图像进行了测试, 这些图像中共包含 29,873 肤色点(LL子带), 表 1 给出了 3 种肤色模型的统计分割结果, 图 3 以Miss America图像为例, 对 3 种模型的分割精度作定性对比, 黑点表示肤色点。肤色分割是在LL子带进行, 所以二值分割图像的分辨率仅是原始图像的 $1/2^k$, k 是小波分解级数。从表 1 和图 3 可以看出, 双适形模型具有更为精确的分割结果, 而其他两种模型误检率相对较高。

肤色分割后的二值图像常常包含一些噪音。这些噪音或者是由于脸部器官如眼睛, 嘴巴的存在而表现出的肤色区域

表 1 3 种模型肤色分割统计结果比较

	正确检测率(%)	误检率(%)
双适形模型	94.34	13.79
适形模型	96.52	33.56
矩形模型	96.68	48.92

注: 表中正确检测率=正确分割出的肤色点数/实际肤色点数; 误检率=非肤色点而误分割为肤色点的数目/实际分割出的肤色点数。

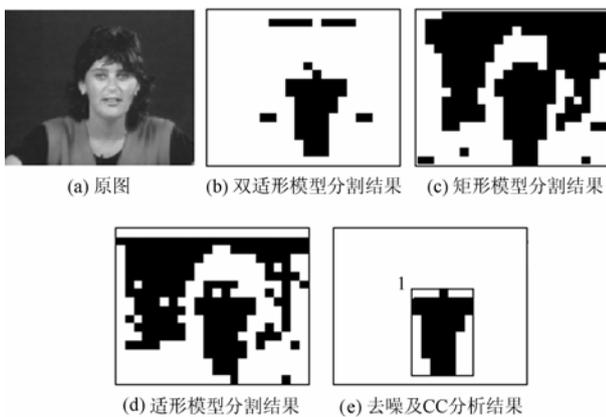


图 3 肤色分割结果示例

中的空穴点, 或者是背景中一些类肤色的孤立点。3×3 模板

的中值滤波器被用来滤除这些噪音。对去噪后的二值图像, 利用连通域分析法对各个独立的连通肤色区标号并记录其位置参数, BAG算法^[9]被用来执行该过程。经连通域分析, 得到一系列目标肤色区: $G = \{r_i(x_i^j, x_i^j, y_i^j, y_i^j)\}$, 其中 i 表示肤色区的标号, $x_i^j, x_i^j, y_i^j, y_i^j$ 是其位置坐标。对图 3(b)的处理结果如图 3(e)所示。

4 基于多级梯度能量描述的人脸检测

对候选肤色区, 采用基于纹理的方法进行搜索, 定位人脸位置和尺寸。

当映射基于纹理的方法至压缩域时, 除提取人脸模式的压缩域视觉纹理描述和选择合适的模式识别方法外, 还需解决人脸尺寸未知(即多尺度人脸检测)的问题^[7,8]。在非压缩域, 人脸尺寸未知的问题常通过固定人脸模板大小, 按一定比率逐级缩小图像来解决。而在压缩域, 任意因子的图像缩放往往较难实现。本文则根据小波压缩数据的特点, 提出人脸模式的多级梯度能量描述, 在高效表征人脸模式的固有特征的同时, 首次较好地解决了人脸尺寸未知的问题。

4.1 多级梯度能量描述的定义

在小波压缩域采用与文献[8]相似的梯度能量来描述人脸特征。相对于脸部肤色区, 脸部器官(眼、鼻、嘴)在图像中表现为边缘跳变, 而梯度能量正好可以反映这种跳变信息, 所以利用梯度能量可突显脸部器官, 从而有效描述人脸^[8]。为避免压缩域图像缩放操作, 本文扩展文献[8]的工作, 提出多级梯度能量描述。多级梯度能量的提出主要是利用小波变换的多分辨率特性, 因不同分辨率(即不同尺度)的细节子图对应原始图像不同频率的边缘跳变信息, 而不同大小的人脸其脸部器官也正好表现为不同频率的边缘跳变, 故用多级梯度能量可更好地描述宽尺寸范围的人脸。

实验中发现, 在小波压缩域, 不同尺寸人脸的脸部器官特征只在部分特定尺度细节子带中表现突出, 也就是不同尺寸人脸的脸部器官在小波系数的不同尺度表现出边缘跳变特性。另外, 在粗尺度细节子带, 小尺寸人脸的脸部特征常常被噪音所遮盖, 无法突显出来。基于这些观察, 设计了人脸模式的多级梯度能量描述, 利用不同尺度的梯度能量来描述不同尺寸的人脸。与文献[8]中仅是针对固定尺寸人脸的单一梯度能量描述不同, 这里提出的针对任意尺寸人脸的多级梯度能量描述, 可避免复杂的压缩域图像缩放操作。

多级梯度能量描述定义如下。

按尺寸将人脸分为 5 个级别: $level_1 \sim level_5$, 分别对应于不同尺寸范围的人脸, 见表 2。由于不同大小人脸的脸部器官在不同尺度的细节子带中高亮突出, 因此对不同级别梯度能量的定义也不同。通常色度分量的高频细节子带携带的信息远小于亮度分量, 故用小波压缩域的Y分量来定义多级

梯度能量即可。

$$G_1(i, j) = \{[\text{LH}_1(i, j)]^2 + [\text{HL}_1(i, j)]^2\} / \text{norm_}c_1 \quad (1)$$

$$G_2(i, j) = \{[\text{LH}_1(i, j)]^2 + [\text{HL}_1(i, j)]^2 + [\text{LH}_2(i/2, j/2)]^2 + [\text{HL}_2(i/2, j/2)]^2\} / \text{norm_}c_2 \quad (2)$$

$$G_3(i, j) = \left\{ \left[\sum_{\substack{m=2i+1 \\ n=2j+1}}^{m=2i+1} (\text{LH}_1(m, n))^2 + (\text{HL}_1(m, m))^2 \right] / 4 + [\text{LH}_2(i, j)]^2 + [\text{HL}_2(i, j)]^2 + [\text{LH}_3(i/2, i/2)]^2 + [\text{HL}_3(i/2, j/2)]^2 \right\} / \text{norm_}c_3 \quad (3)$$

$$G_4(i, j) = \left\{ \left[\sum_{\substack{m=2i+1 \\ n=2j+1}}^{m=2i+1} ((\text{LH}_2(m, n))^2 + (\text{HL}_2(m, m))^2) \right] / 4 + [\text{LH}_3(i, j)]^2 + [\text{HL}_3(i, j)]^2 \right\} / \text{norm_}c_4 \quad (4)$$

$$G_5(i, j) = \{[\text{LH}_3(i, j)]^2 + [\text{HL}_3(i, j)]^2\} / \text{norm_}c_5 \quad (5)$$

其中 $\text{LH}_n(i, j), \text{HL}_n(i, j)$ 分别代表Y分量的 LH_n, HL_n 子带中位置 (i, j) 的小波系数；不同级别梯度能量的定义公式不同，其取值范围也不同，为了简化后续检测器的设计，使不同大小的人脸均可采用同一个检测器进行检测，这里引入用来归一化不同级别梯度能量的参数 $\text{norm_}c_x$ ，首先统计5个级别人脸样本集的梯度能量分布(近似为高斯分布)，然后计算各分布的均值及方差，最后通过调整 $\text{norm_}c_x$ 使得不同级别梯度能量分布基本一致， $\text{norm_}c_x$ 的具体值见表2。从式(1)~(5)可发现，多级梯度能量图 $G_1 \sim G_5$ 的尺寸分别是原始图像的 $1/2, 1/2, 1/4, 1/8$ 和 $1/8$ 。因此，即使待检人脸的尺寸在一个很

如表2所列。故对各种尺寸的人脸，基于梯度能量描述的特征提取计算量基本一致。图4为多级梯度能量描述表征人脸模式的有效性演示，人脸与相应的梯度能量图对应排布，数字标注为其真实尺寸。很明显，多级梯度能量描述总能使人脸固有器官(如眼睛、鼻子、嘴巴)高亮突出，人脸尺寸对其影响极小。

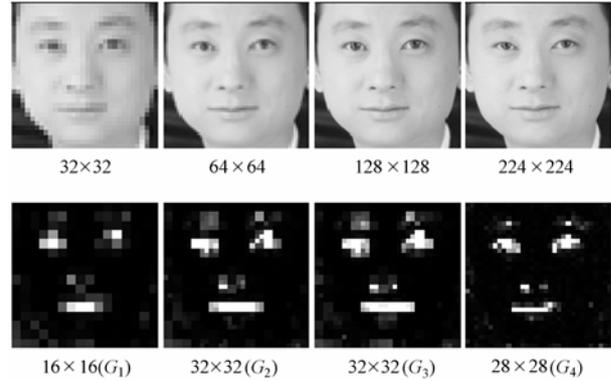


图4 人脸模式的多级梯度能量描述

4.2 人脸模式的梯度能量分布

为分析人脸区的梯度能量分布，本文收集了2301个包含多种光照条件的正面或微斜人脸样本，这些样本是直接来自图像和视频中裁剪眼-鼻-嘴区获得。根据就近原则，所有人脸样本被分组并归一化到不同尺寸(如 $32 \times 32, 40 \times 40, 48 \times 48, \dots, 128 \times 128, \dots$)的集合中。用相应的公式计算各个集合所有样本的梯度能量图并求平均。图5(a)~图5(d)分别是 $32 \times 32, 64 \times 64, 128 \times 128$ 及 160×160 集合的平均梯度能量图。图中脸部器官部位梯度能量分布往往较大，脸颊部位分布则很小。并且由于采用了多级梯度能量描述，对宽尺寸范围人脸，梯度能量分布基本保持一致。

表2 多级梯度能量描述中的相关参数

	Level ₁	Level ₂	Level ₃	Level ₄	Level ₅
人脸尺寸范围 (像素域)	[20×20 ~ 40×40)	[40×40 ~ 80×80)	[80×80 ~ 160×160)	[160×160 ~ 320×320)	[320×320 ~]
相应梯度能量图 尺寸范围	[10×10 ~ 20×20)	[20×20 ~ 40×40)	[20×20 ~ 40×40)	[20×20 ~ 40×40)	[40×40 ~)
涉及到的 细节子带	LH ₁ , HL ₁	LH ₁ , HL ₁ LH ₂ , HL ₂	LH ₁ , HL ₁ LH ₂ , HL ₂ LH ₃ , HL ₃	LH ₂ , HL ₂ LH ₃ , HL ₃	LH ₃ , HL ₃
norm_c	4.0	4.0	3.5	2.0	1.0

大的范围里变动，其梯度能量描述仍具有紧凑的尺寸范围，

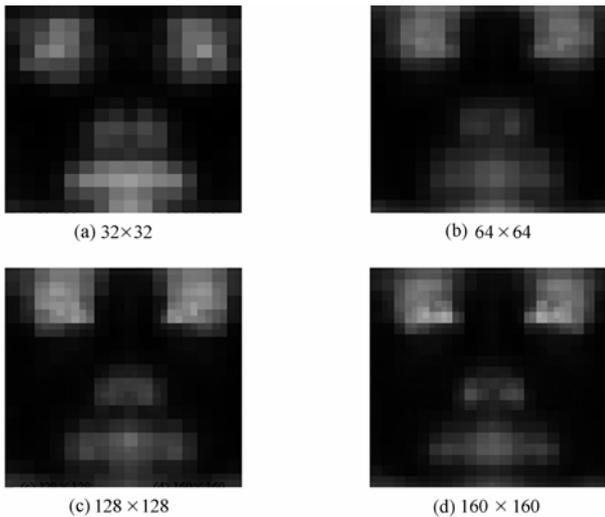


图 5 不同大小人脸区的平均梯度能量分布

4.3 候选肤色区的多级梯度能量描述

在实际操作中, 根据候选区的尺寸计算其中的部分或全部梯度能量图。假定候选区的尺寸为 $R_w \times R_h$, 则该候选区中的最大可能人脸尺寸为 $\min(R_w, R_h) \times \min(R_w, R_h)$ 。若该尺寸落在第 i 级人脸 $level_i$ 见表 2 的尺寸范围中, 就只需计算 $G_1 \sim G_i$ 。如图 3(e) 中的候选肤色区, 其尺寸为 56×80 (在 LL 子带为 7×10), 因 $56 \in [40, 80)$, 故只需计算 G_1 和 G_2 , 而且计算时只涉及到 LH_1, HL_1, LH_2, HL_2 子带中相应区域的小波系数。

获得候选肤色区的多级梯度能量图后, 需判定图中各个尺度各个位置的矩形区域(以下简称检测窗口)是否为人脸, 本文采用简单规则与神经网络相结合的级联检测器来分类每个窗口。

4.4 级联检测器

级联检测器采用与文献[10]提出的层叠分类器相似的结构(见图 6)。本文设计的级联检测器包含 4 个分类器。前 3 个是基于规则的简单分类器, 第 4 个是基于神经网络的复杂分类器。

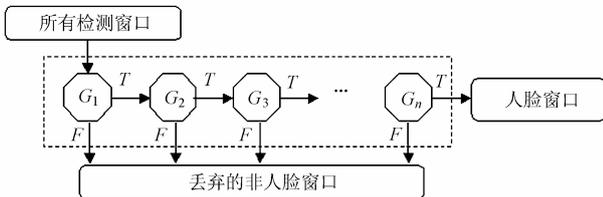


图 6 级联检测器

通过观察图 5, 根据人脸区梯度能量值的分布制定了一些直观简单的规则来逐级确认检测窗口是否存在脸部器官。将检测窗口按图 7(a)~7(c) 所示的模板分区, 计算整个检测窗口及各区的梯度能量和 $E, E_{er}, E_{el}, E_m, E_L, E_R, E_T, E_B$, 分类器 $C_1 \sim C_3$ 按照下列条件逐级判断当前检测窗口是否为人脸, 若

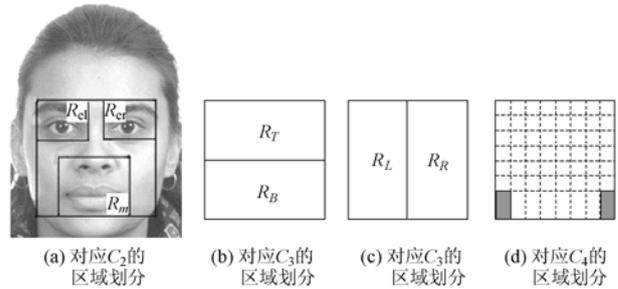


图 7 人脸模板及特征区域划分

满足, 送入下一级分类器继续确认, 否则丢弃。

$$C_1: T_{max} > E > T_{min}$$

$$C_2: E_p / E > T_p, \quad E_p = E_{er} + E_{el} + E_m$$

$$C_3: r_{min}^1 \leq E_T / E_B \leq r_{max}^1 \quad \&\& \quad r_{min}^2 \leq E_R / E_L \leq r_{max}^2$$

对以上 3 个分类器中涉及到的阈值 $T_{min}, T_{max}, T_p, r_{min}^1, r_{max}^1, r_{min}^2, r_{max}^2$, 通过人脸样本集和非人脸样本集训练得到。训练的思路是: 首先计算正样本集中每个样本的 $E, E_{er}, E_{el}, E_m, E_L, E_R, E_T, E_B$ 并统计其分布规律, 根据这些分布规律初始化各阈值, 然后用上述分类器对正负样本集进行分类, 若漏检率几乎为 0 ($< 1\%$), 则调整阈值(如增大 T_{min} , 减小 T_{max}), 以降低误检率, 若漏检率 $> 1\%$, 则调整阈值(如减小 T_{min} , 增大 T_{max}) 降低漏检率。这种调整迭代进行, 直到漏检率几乎为 0, 而误检率也无法再继续降低为止。图 8 给出了调整过程中误检与漏检的关系曲线, 可以看出, 随着误检率的增加, 漏检率不断下降, 最后趋于几乎为 0 的平稳状态。这里, C_1, C_2, C_3 是针对提高检测速度设计的, 其目的是滤除明显的非人脸, 以减轻神经网络分类器 C_4 的负担, 故选择使漏检率几乎为 0, 误检率尽量小的那些阈值(即图 8 中圆圈所示最佳点对应的阈值)作为分类器的最终阈值。非人脸样本集是从多个不包含人脸的图像中采集得来, 共 19,267 个。与决策树类似, 级联分类器的训练也是序贯进行的, 只有那些通过了前面分类器的样本, 才被用来作为后面分类器的训练集。实验中发现, 对测试图像集, C_1 可以滤除大约 40% 的检测窗口, C_2 可进一步滤除 27%, 经过 C_3 后, 只有大约 14% 的检测窗口可进入接下来的神经网络分类器。与只基于用神经网络分类器进行检测的方法相比, 这种级联策略可大大提高检测速度。

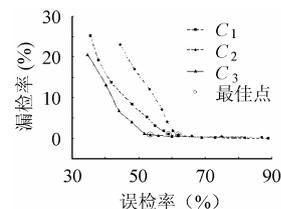


图 8 简单分类器的误检-漏检关系曲线

神经网络可模拟人脑处理复杂事件。因此, 采用基于神

经网络的复杂分类器 C_4 对通过前3个分类器的候选人脸进行确认,进一步滤除非人脸窗口。划分 $M \times M$ 的候选检测窗为64个 $M/8 \times M/8$ 大小子块(见图7(d))。忽略左下角和右下角(图中灰色块)的4个块以消除背景的影响。计算剩余60块的梯度能量平均值 $E_a(i)$, $i=1,2,\dots,60$,作为特征矢量。然后采用两层、前向神经网络分类该特征矢量为人脸还是非人脸。神经网络具有60个输入节点,16个隐藏节点,1个输出节点。用BP算法进行训练,正样本4602个,负样本用“自举法”(boot-strap)^[11]选取,即首先用通过前3个简单分类器的非人脸样本作为负样本集,用这些样本和人脸样本训练神经网络分类器,然后用训练好的检测器(包括1-4级分类器)对大量的非人脸样本进行判决,判决错误的样本作为新的负样本被插入到负样本训练集中。这个过程重复多次即可收集到足够的典型非人脸样本。本文最后共收集到7564个负样本。

4.5 搜索策略及结果仲裁

为解决多尺度人脸的检测问题,本文首次提出缩放检测窗口与上述多级梯度能量描述相结合的搜索策略,即对 $R_w \times R_h$ 大小的候选肤色区,设置初始检测窗口尺寸为 $\min(R_w, R_h) \times \min(R_w, R_h)$,根据表2确定与该尺寸相对应的梯度能量图及搜索窗口尺寸,并在该梯度能量图中以一定步长(对 G_1 取1,其余取2)移动搜索窗口进行检测。以1.2的因子缩小检测窗口,重复上述步骤,直到检测窗口的尺寸小于32, 32×32 是本文方法能检测的最小人脸尺寸。因此,搜索过程只需进行检测窗口的尺寸缩小以及简单的查表操作,避免了复杂的压缩域图像缩放。值得指出的是,这里搜索窗口与检测窗口的分辨率不同,检测窗口的分辨率与原始图像相同,而搜索窗口分辨率与相应的梯度能量图一致,所以对每一轮检测窗口,不但需确定哪一级梯度能量图与其相应,且需映射检测窗口为搜索窗口,以便在对应的梯度能量图中扫描。

因为是在多尺度多位置搜索,所以真正的人脸常在多个相邻的位置及尺度被检测到,而错误报警的一致性相对较差。基于这种现象,本文采用一种既可减少错误报警率又可融合多重检测的结果仲裁方法:即统计每个可能位置邻域内人脸个数,若个数大于某个阈值,人脸判定为真;否则认为是错误报警而丢弃。结果为真的位置,以邻域中所有检测结果的中心和尺度平均值作为最终结果的中心和尺度。在实验结果中将会讨论该阈值对算法性能的影响。

5 实验结果

为验证本文提出算法的有效性和快速性,对大量图像进行了测试。测试图片与训练人脸的图片源不同。根据测试图像的类型及复杂度,分为2个测试集:(1)简单图像测试集,

由头肩图像序列 Miss America 和 Miss AKIYO 中各取101帧组成,其中共包含202个正面大致端正人脸, Miss America 序列背景简单,图中基本无类肤色背景, Miss AKIYO 背景稍复杂,衣服颜色与肤色类似。(2)复杂图像测试集,包括数码相机照片,扫描照片及网络下载图片共296幅,共有正面大致端正各色人脸370个。该测试集中的图像背景复杂,光照变化大,大小不一,大多数存在类肤色背景。所有图片均以JPEG2000压缩格式存放,压缩时采用3级小波分解。

由前面的介绍可知,在肤色模型、简单分类器阈值与神经网络参数都确定的情况下,本文算法的检测性能主要受4.5节中的仲裁阈值的影响,通过调整该阈值,可以得到正确检测率-误检率曲线(测试集2),如图9所示。综合考虑检测率和误检率,本文设置该阈值为3。

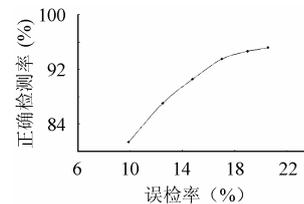


图9 正确检测率-误检率曲线

实验统计数据见表3,以检测率(正确检测人脸数/图像集中包含的人脸数)和误检率(误检人脸数/实际检测出的人脸数)以及平均检测时间作为衡量指标。平均检测时间(括号外数字)包括JPEG2000部分解压(括号中数字)和小波域人脸检测时间两部分,另外,为了说明压缩域处理的优势,表中还给出了全解码时间。图10给出了部分检测结果示例。其中A为Miss AKIYO序列中的图片,B~E为测试集(2)中的图片。

作为比较,表4给出了DCT域典型方法^[8]对2个测试图像集的实验结果,这里图像以JPEG压缩格式存放,平均检测时间(括号外数字)包括部分解压时间(括号中数字)和DCT域人脸检测时间。

从表3,表4可以看出,对于压缩格式图像,本文方法在检测性能和检测速度上均有较大优势。这主要是因为本文提出的多级梯度能量描述有效地避免了文献[8]中复杂的压缩域系数缩放、重组操作,充分体现压缩域处理的优势(即保持检测速度的同时,省去大部分解码时间),大大减少了总的人脸检测时间,而DCT域方法虽然也省去了部分解码时间,但其在DCT系数重组及缩放操作上费时太多,以至于抵消了压缩域处理的优势;同时由于对不同大小人脸根据其特点采用了不同的描述公式,使得对人脸模式的描述更加客观,从而较大地改善了检测性能。另外,肤色信息的合理使用、级联检测器的逐级淘汰机制也在一定程度上提高了本文算法的速度。

表3 本文方法测试实验统计结果

测试集	正确检测数	漏检数	误检数	检测率(%)	误检率(%)	平均检测时间(s)	全解压平均时间(s)
简单测试集	200	2	8	99.01	3.85	0.076(0.023)	0.078
复杂测试集	346	24	71	93.51	17.03	0.359(0.087)	0.241

表 4 DCT域方法^[8]实验统计结果

测试集	正确检测数	漏检数	误检数	检测率(%)	误检率(%)	平均检测时间(s)	全解压平均时间(s)
简单测试集	199	3	10	98.51	4.78	0.124(0.021)	0.073
复杂测试集	310	60	72	83.78	18.85	1.196(0.075)	0.198



图 10 部分检测结果

6 结束语

本文针对 JPEG2000 压缩格式彩色图像, 提出了一种结合肤色和纹理信息的人脸检测方法。该方法的最大贡献是充分利用小波数据的特点, 提出多级梯度能量描述和缩放检测窗口尺寸相结合的搜索策略, 首次较好地解决了压缩域人脸检测中尺寸未知的问题, 实现了小波压缩域多尺度、多位置人脸的快速检测。此外, 算法将基于规则的简单分类器与基于神经网络的复杂分类器有机结合, 在保证检测正确率的基础上, 大大提高了处理速度。试验结果表明, 这种基于压缩域的人脸检测方法是高效可行的。

参考文献

- [1] 梁路宏, 艾海舟, 徐光佑等. 人脸检测研究综述. 计算机学报, 2002, 25(5): 449 - 458.
- [2] Chai D, Ngan K N. Face segmentation using skin-color map in videophone applications. *IEEE Trans. on Circuits and Systems for Video Technology*, 1999, 9(4): 551 - 564.
- [3] Garcia C, Tziritas G. Face detection using quantized skin color regions merging and wavelet packet analysis. *IEEE Trans. on Multimedia*, 1999, 1(3): 264 - 277.
- [4] Rowley H A, Baluja S, Kanade T. Neural network-based face

detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998, 20(1): 23 - 38.

- [5] Osuna E, Freund R, Girosi F. Training support vector machines: an application to face detection. In: Proc IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Juan, 1997: 130 - 136.
- [6] Wang H L, Chang S F. A highly efficient system for automatic face region detection in MPEG video. *IEEE Trans. on Circuits and Systems for Video Technology*, 1997, 7(4): 615 - 628.
- [7] Luo H T, Eleftheriadis A. On face detection in the compressed domain. In: Proceedings of the Eighth ACM International Conference on Multimedia, California, 2000: 285 - 294.
- [8] Chua T S, Zhao Y L, Kankanhalli M S. Detection of human faces in a compressed domain for video stratification. *The Visual Computer*, 2002, 18: 121 - 133.
- [9] Bin Y, Jain A K. A generic system for form dropout. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1996, 18(11): 1127 - 1134.
- [10] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Hawaii, 2001, 1: 511 - 518.
- [11] Sung K K, Poggio T. Example-based learning for view-based human face detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1998, 20(1): 39 - 51.

李晓华: 女, 1973 年生, 讲师, 研究方向为人脸检测、基于内容图像检索、图像压缩等。

沈兰荪: 男, 1938 年生, 教授, 博士生导师, 研究领域为图像处理、视频编码与传输、医学图像处理等。