

# 一种基于多级梯度能量特征的 DCT 压缩域人脸检测算法

李晓光, 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:** 压缩域人脸检测在图像/视频信息处理中具有重要意义. 本文提出了一种基于多级梯度能量特征的 DCT (Discrete Cosine Transform) 压缩域人脸检测算法. 依据 DCT 压缩图像色差信号的直流系数进行肤色分割, 减小检测范围. 在分割为肤色的区域提取多级梯度能量特征, 即利用不同大小的检测窗口提取归一化的特征向量, 表示不同大小的人脸. 特征向量输入到级联分类器中分类, 确定是否表示人脸. 级联分类器由若干简单分类器和一个神经网络分类器构成. 简单分类器利用一些先验知识排除大部分明显不是人脸的特征向量, 通过简单分类器的特征由神经网络最终确定是否表示人脸. 多级梯度能量特征与 DCT 域图像缩放相结合实现了对不同大小人脸的快速检测. 对多级梯度能量特征的定义, 减少了检测算法中压缩域图像缩放的次数, 从而大幅度减少了计算复杂度, 提高了检测速度. 实验结果表明提出的多级梯度能量特征可有效描述 DCT 域人脸模式, 同时也证明了该算法的快速有效性.

**关键词:** 人脸检测; DCT; 压缩域; 多级梯度能量

中图分类号: TP391 文献标识码: A 文章编号: 0372-2112 (2005) 12-2170-04

## An Algorithm of Multilevel Gradient Energy based Face Detection in DCT Compressed Domain

LI Xiao-guang, LI Xiao-hua, SHEN Lan-sun

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

**Abstract:** Face detection is important in the processing of images and video. Based on multilevel gradient energy (MGE), an algorithm of face detection in DCT (Discrete Cosine Transform) compressed domain is presented. In preprocessing procedure, skin color segmentation based on the DC of chromatic components is applied to the input image for reducing the detected regions. According to the map of skin segmentation, MGE based feature vector is extracted, viz. normalized feature vectors are extracted from the detecting windows of various sizes to describe faces of different sizes. Then cascade classifier is employed to classify the feature vectors as face or non-face. Cascade classifier is comprised of several simple classifiers and a neural network classifier. Lots of feature vectors that belong to non-face are removed by simple classifiers which embedded preknowledge rules. The left vectors are classified by neural network. We combined MGE features together with image scaling to allow faces of various sizes. The simplicity of feature extraction accelerated detection by reducing the times of image scaling which is more time cost. The experiment results show that the proposed method is efficient and effective.

**Key words:** face detection; DCT; compressed domain; multilevel gradient energy

## 1 引言

随着信息技术的发展, 以图像/视频为主体的多媒体信息日益普及, 逐渐深入地影响着社会的发展和人们生活质量的提高. 人脸检测技术在对图像/视频信息处理中具有重要的意义<sup>[1]</sup>. 目前海量的图像/视频信息以压缩格式存储着, 大多数压缩标准如 JPEG, MPEG, H. 26L 等都是基于 DCT 变换为基础的. 传统的人脸检测方法是在像素域进行的, 要对图像/视频进行解压缩、检测甚至再压缩的环节, 在实际应用中受到限制. 压缩域处理以及人脸检测逐渐成为研究热点<sup>[2-5]</sup>.

压缩域人脸检测的研究开始于 20 世纪 90 年代末, Wang 与 Chang<sup>[6]</sup>以肤色为基础在选定的色度空间建立肤色的统计

模型, 提出了针对 MPEG 视频序列的人脸区域检测方法. 这种方法设计简单, 速度快, 但是背景接近肤色时, 误报比较高; Luo 和 Efetheriadis<sup>[7]</sup>通过 DCT 系数获得颜色与纹理信息, 提出了基于多模高斯聚类的方法来检测人脸, 但是检测效果不够理想, 而且时间复杂度也较高; Chua 和 Zhao<sup>[8]</sup>利用梯度能量特征能够很好的表征人脸, 但是需要进行多次 DCT 域图像缩放及插值操作, 速度上难以令人满意.

本文提出了一种基于多级梯度能量特征与肤色信息相结合的 DCT 人脸检测方法, 避免了过多的 DCT 域图像缩放以及为了获取梯度能量而进行的 DCT 系数重组操作, 从原理上降低了人脸检测算法的时间复杂度.

## 2 算法整体框架

算法的整体框架如图 1 所示. 人脸检测算法处理的对象为 JPEG 等压缩格式图像经过部分解压缩得到的 DCT 数据. 实验中采用 BMP 图像经 DCT 变换得到图像的 Y Cb Cr 三分量的 DCT 数据. 检测经肤色分割、梯度能量特征提取和级联分类器分类, 最后经过后处理以矩形方框的形式输出检测结果.

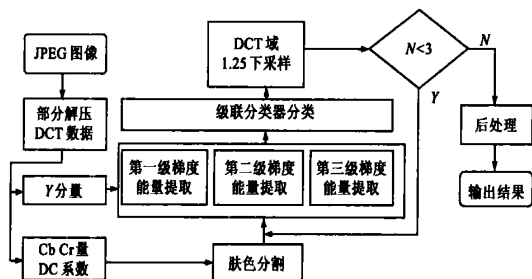


图 1 算法整体框架

肤色分割是利用 DCT 系数 Cb Cr 的直流分量进行的, 通过简单的二维椭圆肤色模型<sup>[9]</sup>分割生成黑白二值图像, 即 skir map. Skir map 的作用是指出下一步梯度能量的提取位置, 使得梯度能量图的提取只在肤色区域进行, 以此来提高特征提取的速度.

现有的一些方法为了检测到不同大小的人脸, 往往需要对原图像进行多次缩放, 在不同尺度下进行多次检测. 但是缩放会引入大量的矩阵计算, 即使有一些快速算法<sup>[10]</sup>, 但是实际效果仍然难以令人满意.

本文算法为人脸特征的提取定义了三级梯度能量, 分别对应三种不同大小的人脸. 在不进行图像缩小的情况下, 分别对应  $32 \times 32$ ,  $56 \times 56$ ,  $112 \times 112$  的人脸. 这样在同一尺度的图像中可以检测三种尺寸的人脸, 从而减少了原图像的 DCT 域缩放操作次数, 提高了检测速度. 算法只须进行三次缩放操作, 便可以检测从  $32 \times 32$  到  $220 \times 220$  之间的人脸.

三级梯度能量特征经过量化构成特征向量, 由级联分类器进行人脸和非人脸的分类. 级联分类器由若干个基于人脸几何特征的弱分类器和一个三层 BP 人工神经网络分类器组成. 经过训练, 达到分类性能的要求.

最后经过后处理输出分类结果.

## 3 算法描述

### 3.1 肤色模型与肤色分割

算法采用的肤色空间为 Cb Cr 空间, 因为研究表明肤色信息在 YCbCr 空间比较集中, 而色度能量主要集中在 Cb Cr 分量上. 肤色模型采用二维椭圆肤色模型<sup>[9]</sup>:

$$\frac{(x - e_x)^2}{a^2} + \frac{(y - e_y)^2}{b^2} \leq 1 \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b - c_x \\ C_r - c_y \end{bmatrix} \quad (1)$$

其中  $c_x = 109.38$ ,  $c_y = 152.02$ ,  $e_x = 1.6$ ,  $e_y = 2.41$ ,  $a = 25.39$ ,  $b = 14.03$  由  $(C_b, C_r)$  计算出  $(x, y)$ , 然后根据椭圆模型公式判断该点是否为肤色点, DCT 域的肤色分割是采用 Cb Cr

分量的直流 DCT 系数进行的. 每个直流系数对表示其所在  $8 \times 8$  DCT 块的平均颜色信息, 因此其分割结果代表了对应  $8 \times 8$  像素块是否为肤色. 分割结果如图 2 所示, (c) 是经过中值滤波去除噪声后的 skir map, 其中每个像素对应原图像中一个  $8 \times 8$  的像素块.



图 2 DCT 域肤色分割以及 SKIN-MAP

### 3.2 多级梯度能量特征

参考文献<sup>[11]</sup>的思想, 定义了三级梯度能量特征来描述大小不同的人脸. 梯度能量在 DCT 域人脸检测中得到了一定的重视. 在文献<sup>[8]</sup>的工作中, 对梯度能量进行了如下的定义:

$$E_{V1}(k) = \sum_{m,n} [DCT(m,n)]^2, m,n \in V \quad (2)$$

$$E_{H1}(k) = \sum_{m,n} [DCT(m,n)]^2, m,n \in H \quad (3)$$

图 3 给出了公式中  $V, H$  区域的定义,  $E_{V1}$  和  $E_{H1}$  分别代表了第  $k$  个 DCT 块的垂直梯度能量和水平梯度能量.

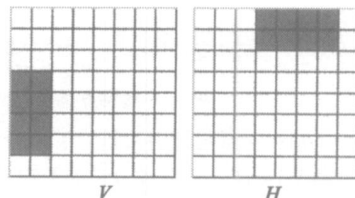


图 3 梯度能量的定义

第一级梯度能量特征针对  $32 \times 32$  的人脸, 直接采用文献<sup>[8]</sup>中的方法: 用  $8 \times 8$  的方窗在图像的 Y 分量 DCT 系数上平移, 横向纵向步长均取 4 个像素 (图 4(a)). 当方窗与原  $8 \times 8$  DCT 块重合

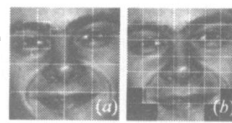


图 4 第一级梯度能量特征的提取

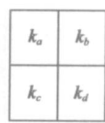


图 5 第三级梯度能量量定义

时按公式 (2) ~ (3) 提取纵向与横向梯度能量. 当方窗与原  $8 \times 8$  DCT 块不重合时, 需要对 DCT 系数按文献<sup>[12]</sup>中的方法进行重新组合, 再提取梯度能量. 从而得到由 49 个梯度能量值描述的人脸梯度能量图. 为了避免脸颊区域背景的影响, 选取图 4(b) 黑色位置以外的 43 个位置, 分别提取和构成 86 维特征向量.

第二级梯度能量特征, 针对  $56 \times 56$  的人脸, 用  $8 \times 8$  的方窗以 8 个像素步长平移, 按公式 (4)、(5) 定义梯度能量:

$$E_{V2}(k) = \alpha \times \sum_{m,n} [DCT(m,n)]^2, m,n \in V \quad (4)$$

$$E_{H2}(k) = \alpha \times \sum_{m,n} [DCT(m,n)]^2, m,n \in H \quad (5)$$

其中  $\alpha = 3.2$  为归一化系数.

同样提取出 86 维特征向量表示人脸特征. 这样不需要 DCT 系数重新组合, 降低了特征提取的计算量.

第三级梯度能量特征的提取, 针对  $112 \times 112$  大小的人

脸,按公式(6)、(7)定义梯度能量:

$$E_{V2}(k)=\alpha\times(E_{V1}(k_a)+E_{V1}(k_b)+E_{V1}(k_c)+E_{V1}(k_d))\quad(6)$$
$$E_{H2}(k)=\alpha\times(E_{H1}(k_a)+E_{H1}(k_b)+E_{H1}(k_c)+E_{H1}(k_d))\quad(7)$$

其中 $\alpha=1$ 为归一化系数.

在16×16像素范围内提取第三级梯度能量,如图5所示,将其中包含的四个8×8DCT块的第一级梯度能量相加,乘以归一化系数.运算过程无须DCT系数重组.按以上相同的方法提取86维特征向量表征112×112大小的人脸.

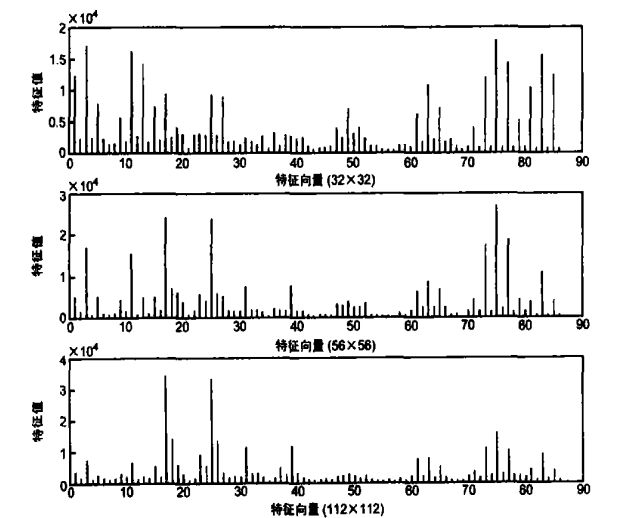


图6 三级梯度能量特征(未归一化)

归一化系数的引入是为了消除三级梯度能量特征的取值范围差异,从而使后续分类器的输入范围保持一致,图6所示是未归一化的三级梯度能量特征向量,可以直观地看到三级向量的相似性,但其取值范围明显不同.

检测过程是对图像的每个缩放尺度分别用32×32,56×56,112×112的窗口进行步长分别为4,8,8平移,按三级能量梯度的定义提取特征向量进行分类.由于特征向量的取值范围比较大,在送入分类器之前需要按图7的曲线<sup>[8]</sup>进行量化.这样做可以把输入向量范围映射到[0,50]区间内,避免出现特别大的数值影响分类器的性能.

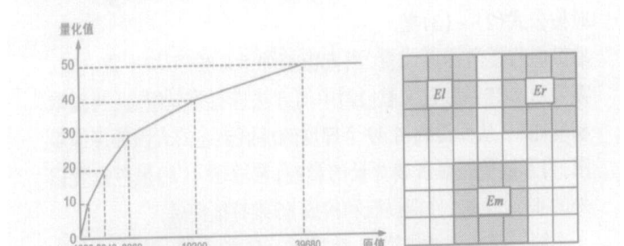


图7 梯度能量量化曲线

图8 C2的区域定义

表1 三级能量梯度特征人脸尺度的范围

特征	不缩小	1.25倍 缩小1次	1.25倍 缩小2次	1.25倍 缩小3次
第一级特征	32×32	40×40	50×50	64×64
第二级特征	56×56	70×70	88×88	110×110
第三级特征	112×112	140×140	176×176	220×220

须进行三次DCT域1.25倍缩小操作<sup>[13]</sup>,便可以检测到大小范围为32×32到220×220范围的人脸,具体如表1所示.其中灰色标记的尺寸与其他的比较接近,因此检测过程中可以不考虑.

3.3 级联分类器

算法采用级联分类器对特征向量进行分类.由两个弱分类器C1、C2和一个BP人工神经网络构成.设 $f=(x_1,x_2,\dots,x_{86})$ 为量化后的特征向量,则分类器描述为:

$$C1: \text{if } \sum_{i=1}^{86} x_i < th1 \text{ then } f \in \text{non\_face};$$
$$C1: \text{else if } \sum_{i \in El+Er+Em} x_i \setminus \sum_{i=1}^{86} x_i < th2 \text{ then } f \in \text{non\_face}; \quad (14)$$

BPNN: eles BPNN( $f$ )

其中th1,th2是由训练样本确定的阈值.El,Er,Em区域如图8所示.在梯度能量特征中,人脸的眼睛,鼻子和嘴三个区域的梯度能量较大,因此C1排除一些梯度能量之和过小的特征,即所对应的区域过于平坦.C2是基于人脸的几何结构,人脸的眼睛和嘴鼻区域的梯度能量之和与总梯度能量和的比应该大于阈值.以此排除大部分明显的非人脸,以减少神经网络分类器的工作量,提高分类速度.

通过前两个弱分类器的特征,由训练好的神经BP网络分类.

神经网络分类器的结构为3层的BP神经网络,86个输入节点,16个隐含节点,1个输出节点.训练神经网络采用自举的方法进行.采集正样本2048个,负样本随着训练过程逐渐采集.采用MATLAB神经网络工具箱进行训练.

4 实验结果以及分析

为了验证算法的性能,采用C++编程,在P4 2.4GHz的PC机上进行实验.测试集为包括数码相机拍摄的照片,扫描普通照片及网上下载的图片,共286幅,共有正面大致端正各色人脸360个.该测试集中的图像背景复杂,光照变化较大,大小不一,大多数存在类肤色背景.

实验比较了两种方案,方案一为本文提出的算法.方案二是文献[8]中方法的近似实现,即采用单一的梯度能量特征通过不断的DCT域缩放检测,并将肤色信息用在后面的结果验证.检测结果见表2.算法速度实验结果见表3.

表2 实验结果比较

算法	总人脸数	正确检测	漏检	正确检测率	误检率
方案1	360	294	66	81.7	26.1%
方案2	360	301	59	83.6	30.4%

①正确检测率= 正确检测人脸/ 总人脸数×100%

②误检率= 误检结果/ 产生结果总数×100%

表3 算法耗时比较(单位:s)

算法	总时间	特征提取时间	分类时间	DCT缩放时间	后处理时间
方案1	177.470	79.181	18.611	64.522	8.762
方案2	342.254	217.136	31.623	79.137	8.573
速度提高比	48%	64%	41%	18%	

通过三级梯度能量特征的定义来检测人脸,对原图像只

表中列出时间为处理 286 幅图像集合的总处理时间, 列出的为主要耗时部分的时间, 因此各项之和不等于总时间。

从实验结果上看, 两种方案的正确检测结果相差不大, 可以采取合理的后处理验证方法进一步提高正确检测率。方案一的检测率略有下降, 这是因为两种方案采用同一个神经网络分类器, 训练样本均选用第一级梯度能量特征, 更适合方案二的算法。

表 2 中的实验结果上说明本文提出多级梯度能量特征可以有效地描述 DCT 压缩域的人脸模式。从算法的速度上看, 多级梯度能量的优势是很明显的。表 3 的实验结果显示处理整个测试集合的总时间提高了 48%, 其中特征提取部分的时间提高的幅度最大, 达到 64%。附带的 DCT 缩放以及检测速度都有相应的提高。

需要说明的是后处理时间主要取决于多个候选位置的融合过程, 分类器的输出是影响后处理的速度主要因素。方案二中的肤色验证速度很快, 处理整个图像库 287 幅图像仅需要 0.06s, 因此出现了方案一的时间稍长的结果。

## 5 结束语

压缩域人脸检测技术在多媒体信息处理领域具有广泛的应用前景。本文提出了基于多级梯度能量特征的 DCT 压缩域人脸检测算法, 详细介绍了算法实现过程。算法从原理上提高了特征提取的速度。实验证明这种多级梯度能量特征能够很好的描述 DCT 域的人脸模式。

压缩域人脸检测的研究是一项新工作, 还有很多问题值得深入研究。本文算法还存在一些有待完善的地方。比如误检率比较高。在今后的工作中, 拟考虑建立适合 DCT 域的肤色模型以及利用更多的人脸模式中其他的一些特征来进一步提高算法的正确检测率, 降低误检率。

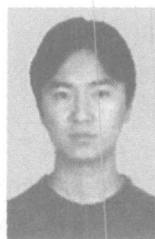
## 参考文献:

- [1] 沈兰荪, 等. 视频编码与低速率传输[M]. 北京: 电子工业出版社, 2001. 134- 167.  
Shen L S, et al. Video coding and low bit rate transmission [M]. Beijing: Publishing House of Electronics Industry, 2001. 134- 167. (in Chinese)
- [2] 沈兰荪. 压缩域图像/视频处理技术的研究[J]. 计算机自动测量与控制, 2000, 8(5): 1- 3.  
Shen L S. Research on compressed domain image/video processing Techniques[J]. Computer Automated Measurement & Control, 2000, 8(5): 1- 3.
- [3] 沈兰荪, 李晓华, 黄祥林. 压缩域图像检索技术研究[J]. 北京工业大学学报, 2003, 29(4): 385- 390.  
Shen L S, Li X H, Huang X L. Research on compressed domain image retrieval technique[J]. Journal of Beijing University of Technology, 2003, 29(4): 385- 390. (in Chinese)
- [4] 李晓华, 沈兰荪. 基于小波压缩域的统计纹理特征提取方法[J]. 电子学报, 2003, 31(12A): 2123- 2126.  
Li X H, Shen L S. Statistical texture feature extraction based

on wavelet compressed domain[J]. Acta Electronica Sinica, 2002, 32(12A): 2123- 2126. (in Chinese)

- [5] 田巍, 黄祥林, 沈兰荪. 一种用于 DCT 压缩域的人脸检测算法[J]. 测控技术, 2004, 23(5): 41- 43.  
Tian W, Huang X L. An algorithm of face detection in the DCT compressed domain[J]. Measurement & control technology, 2004, 23(5): 41- 43. (in Chinese)
- [6] Wang H, Chang S F. A highly efficient system for automatic face region detection MPEG video[J]. IEEE Trans on Circuit System Video Technol, 1997, 7(4): 615- 628.
- [7] Lou H, Eleftheriadis A. On face detection in the compressed domain[A]. ACM Multimedia Conference[C]. Los Angeles: ACM Press, 2000. 285- 294.
- [8] T S Chua, Y L Zhao, M S Kankanhalli. Detection of human faces in a compressed domain for Video stratification[J]. The Visual Computer. 2002, 18: 121- 133.
- [9] Jure Kovac, Peter Peer, Franc Solina. 2D versus 3D colour space face detection[R]. Zagreb: 4th EURASIP Conference focused on Video/ Image and Multimedia Communications, 2003, 2- 5.
- [10] Neri Merhav, Vasudev Bhaskaran. Fast algorithms for DCT-domain image downsample and for inverse motion compensation[J]. IEEE Trans on Circuits and Syst for Video Technol. 1997, 7(3): 468- 476.
- [11] 李晓华. 小波压缩域图像检索技术的初步研究[D]. 北京: 北京工业大学博士论文, 2004.  
Li X H. Research on image retrieval techniques in the wavelet compressed domain[D]. Beijing: Beijing University of Technology, 2004. (in Chinese)
- [12] Shih Fu Chang, David G. Messerschmitt. Manipulation and composition of MG DCT compressed video[J]. IEEE Journal on selected areas in communication, 1995, 13(1): 1- 11.
- [13] Y L Zhao, M S Kankanhalli, T S Chua. Fractional Scaling of Image and Video in DCT domain[A]. ICIP03[C]. Barcelona: ICIP, 2003, (1): 185- 188.

## 作者简介:



李晓光 男, 1980 年 8 月出生于北京, 2003 年毕业于北京工业大学, 获学士学位, 现为北京工业大学信号与信息处理研究室硕士研究生, 主要研究方向为图像处理技术。  
E-mail: lxg@emails.bjut.edu.cn.

李晓华 女, 1973 年 2 月出生于陕西韩城, 讲师, 博士, 主要研究方向为压缩域图像处理、图像压缩、图像理解等。

沈兰荪 男, 1938 年 6 月出生于江苏徐州, 教授、博士生导师, 研究领域为图像编码、图像处理、传输与应用以及光谱信号的检测与处理。