

## 基于 DCT 压缩域的纹理图像分类<sup>1</sup>

黄祥林 沈兰荪

(北京工业大学信号与信息处理研究室 北京 100022)

**摘 要** 该文提出了一种在基于离散余弦变换 (DCT, Discrete Cosine Transform) 压缩域进行纹理图像分类的方法。此方法主要利用了 DCT 域能量具有方向性的特点, 直接在游程长编码 (RLE, Run Length Encoding) 形成的码流中根据这种方向性来提取、组织特征矢量, 对纹理图像进行分类。此方法只需对压缩图像进行简单的 Huffman 解码, 要处理的数据量较少, 处理速度快。实验结果表明, 这种分类方法具有较高的准确性。

**关键词** 压缩域图像处理, DCT, 纹理图像分类, 能量方向性

**中图分类号** TN911.73

### 1 引 言

由人类视觉感知心理的研究可知, 视觉纹理特性主要有: 粗糙度、对比度、方向度、线象度、规整度、粗略度等<sup>[1]</sup>, 这些特点是对纹理图像进行处理分析的依据。本文设计了一种基于 DCT 压缩域的纹理图像分类的新方法, 主要出于以下考虑: (1) 无论是网上传输还是本地存储的图像数据都是以压缩的形式出现的, 压缩图像数据具有普遍性; (2) 现有的压缩编码国际标准几乎都是以 DCT 为基础的, 研究基于 DCT 压缩域的纹理图像分类具有代表性; (3) 与传统的分类算法相比, 此方法直接处理 RLE 码流数据, 对于以 DCT 为编码基础的压缩图像只需进行简单的 Huffman 解码, 处理的数据量较少、节省了处理时间; (4) 本文提出的 DCT 域能量具有方向性的特点具有分类纹理图像的能力, 这是本文设计压缩域分类方法的基础。

实际上, 基于压缩域的图像分析处理技术是近年来兴起的多媒体研究领域的一个热点, 一些学者已经提出了许多基于压缩域的多媒体数据分析处理算法<sup>[2-6]</sup>。Shih-Fu Chang<sup>[1]</sup>对 DCT 压缩域的纹理分类进行了研究, 他把 DCT 的每个通道能量作为纹理分类的特征 (对于  $8 \times 8$  DCT, 共有 64 个特征) 来分类纹理图像, 这种特征组织方法并未明显表征纹理图像的特点, 抗干扰能力也不是很强。本文对 DCT 的每个通道能量进行合理的组合, 提出了 DCT 区域能量具有方向性的特点, 这种特点适合对纹理图像进行分类。下面首先介绍 DCT 域的重要特性, 并根据这种特性设计基于 DCT 压缩域的纹理图像分类方法, 最后给出实验结果和结论。

### 2 DCT 域的重要特性

以 DCT 变换为基础的压缩编码算法的基本流程如图 1, 对输入的原始图像进行  $8 \times 8$  DCT 变换、量化、Zigzag 排序、RLE 编码、Huffman 编码, 最终形成压缩码流。

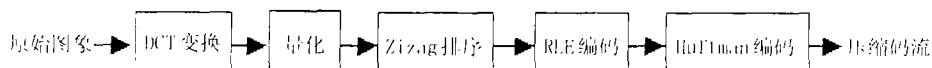


图 1 基于 DCT 变换的压缩编码算法的基本编码框图

<sup>1</sup> 2000-07-28 收到, 2001-01-19 定稿

863-317 计划资助 (863-317-01-10-99), 国家自然科学基金 (60172045) 资助

Huffman 编码形成的码流中数据项不具备明显的结构特点, 其位置也不是字节对齐, 不太利于计算机的数据分析处理, 故在压缩域处理数据时, 一般先对压缩码流进行 Huffman 解码, 得到 RLE 的码流数据, 再在此基础上进行基于 DCT 压缩域的图像分析处理。从经验得知<sup>[3]</sup>, 使用 DCT 编码方法来压缩图像, 当压缩比为 24:1 时, 图像质量不会出现严重失真, 压缩比为 10:1 时, 人眼基本上辨别不出图像的失真, 而 Huffman 编码的压缩比一般为 2.5:1, 因此, 即使对压缩数据进行了 Huffman 解码, 要处理的数据量依然很少 (当图像压缩比为 25:1 时, 经 DCT 变换、量化后 RLE 的压缩比为 10:1)。

DCT 的特性很多<sup>[7]</sup>, 在图像编码应用中它的最主要特性是: 原始图像经 DCT 变换后, 能量主要集中在左上角区域 (即直流、低频和中频区), 而右下角区域 (即高频区) 的能量分布比较弱, 量化后会出现许多零值。正是 DCT 的这种特性, 才具有压缩图像数据的能力。

DCT 的另一个重要特性是它的能量具有方向性, 即图像数据经 DCT 变换后, 其一定区域的变换系数代表着一定方向的频谱成分。如图 2(a) 所示, 区域  $\Omega_0$  表示的是直流成分 (实际上只有左上角的一个系数点), 区域  $\Omega_1$  主要代表了原图像中水平方向的频率变化 (竖直纹理), 而区域  $\Omega_2$  则主要集中了原图像中对角方向的频率变化 (对角纹理), 区域  $\Omega_3$  基本上是原图像中垂直方向的频率变化 (水平纹理)。图 2(b) 是这 4 个区域的 DCT 系数位置的模板矩阵  $M$ 。这里用  $P_0$  表示直流能量,  $P_k (k = 1, 2, 3)$  分别代表不同纹理方向的区域能量,  $N$  表示图像块中 DCT 块的数目。用一般的能量表达式表达为

$$P_0 = \sum_{m=1}^N (A_{00}^m)^2 \quad (1)$$

$$P_k = \sum_{m=1}^N \sum_{\Omega_k} (A_{ij}^m)^2, \quad k = 1, 2, 3 \quad (2)$$

其中  $A_{ij}^m$  表示第  $m$  个 DCT 块中的系数 ( $i, j \in \Omega_k$ )。

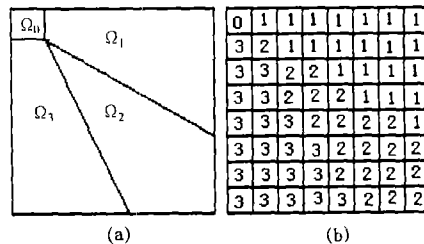


图 2 DCT 区域能量分布  
(a) DCT 域的 4 个区域能量分布 (b) 相对应系数位置的模板矩阵  $M$

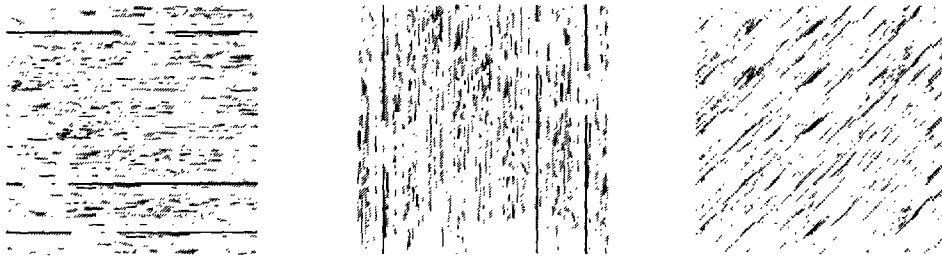
但是, 如果直接用 (1), (2) 式定义的能量来表达图像的纹理特征, 实验结果并不好, 因为输入的任何变化都能引起各区域能量的较大变化, 而各区域能量的变化并不一致。因此, 作者使用了下面的能量表达式:

$$E_0 = \sum_{m=1}^N (A_{00}^m - \bar{A}_0)^2 \quad (3)$$

$$E_k = \sum_{m=1}^N \sum_{\Omega_k} (A_{ij}^m - \bar{A}_k)^2, \quad k = 1, 2, 3 \quad (4)$$

其中  $\bar{A}_0 = \frac{1}{N} \sum_{m=1}^N A_{00}^m$ ,  $\bar{A}_k = \frac{1}{21N} \sum_{m=1}^N \sum_{\Omega_k} A_{ij}^m$ ,  $k = 1, 2, 3$ ;  $\Omega_k$  中的系数个数为 21。实验证明, 用能量  $E_0$ ,  $E_1$ ,  $E_2$ ,  $E_3$  作为纹理的分类特征, 能较好地表示纹理图像的方向模式<sup>[8]</sup>。

图 3(a), 3(b), 3(c) 分别为水平、垂直、对角纹理图像 (取自 Adobe 纹理图像库中的 lines1, lines2, lines3)。表 1 是这些图像经 DCT 变换后各区域能量的分布情况。可见, 对于垂直纹理有  $E_1 > E_2$ ,  $E_1 > E_3$  (水平方向的频率变化); 对于对角纹理有  $E_2 > E_1$ ,  $E_2 > E_3$ , 且  $E_1$  与  $E_3$  相差不大 (对角方向的频率变化); 对于水平纹理有  $E_3 > E_1$ ,  $E_3 > E_2$  (竖直方向的频率变化)。这充分说明了 DCT 域中区域能量具有方向性的特点, 下面就根据 DCT 域的这一特性来设计在压缩域的纹理分类方法。



(a) 水平纹理 (lines1)

(b) 垂直纹理 (lines2)

(c) 对角纹理 (lines3)

图 3 三种典型纹理

表 1 各纹理图像的区域能量值

纹理图像	能量 $E_0$	能量 $E_1$	能量 $E_2$	能量 $E_3$
水平纹理 (a)	88.01	12.81	7.16	151.56
垂直纹理 (b)	86.97	150.61	7.23	12.86
对角纹理 (c)	61.19	45.36	50.86	46.97

### 3 基于 DCT 变换压缩域的纹理图像分类算法设计

#### 3.1 分类特征的选择

作者使用 DCT 域的具有方向性的区域能量作为纹理图像的分类特征。因为人们对纹理的感知主要是纹理的方向和尺度, 而从上面的分析可知, DCT 域的区域能量具有方向性。

#### 3.2 特征矢量维的选择

把 DCT 区域能量作为纹理图像的分类特征, 采用  $4 \times 4$  维 (即 4 组区域能量, 每组区域能量有 4 个区域的能量值) 的特征矢量来分类纹理图像。这里设计了组织特征矢量的两种方法: 第 1 种方法是将整个图像的 DCT 域分成 4 组  $Z_1, Z_2, Z_3, Z_4$ , 每组中有 4 个区域能量  $Z_i = \{E_0^{(i)}, E_1^{(i)}, E_2^{(i)}, E_3^{(i)}\}$ ,  $i = 1, 2, 3, 4$ 。这样就组成了 16 维的特征矢量  $Z = \{Z_1, Z_2, Z_3, Z_4\}$ , 如图 4(a)。这种特征矢量强调的是 4 个局部区域的细节, 缺乏对整幅图像特征的刻画, 故此, 设计了第 2 种组织特征矢量的方法: 第 1 组是整个 DCT 域的区域能量  $Z_1 = \{E_0^{(1)}, E_1^{(1)}, E_2^{(1)}, E_3^{(1)}\}$ , 第 2 组是对整个 DCT 域的  $1/2$  抽样 (以  $8 \times 8$  的 DCT 块为单位) 形成的区域能量  $Z_2 = \{E_0^{(2)}, E_1^{(2)}, E_2^{(2)}, E_3^{(2)}\}$ , 同样, 第 3 组、第 4 组是对整个 DCT 域的  $1/4, 1/8$  抽样形成的区域能量  $Z_i = \{E_0^{(i)}, E_1^{(i)}, E_2^{(i)}, E_3^{(i)}\}$ ,  $i = 3, 4$ , 如图 4 (b), 这种组织方式从一定程度上体现了多分辨率的特征。

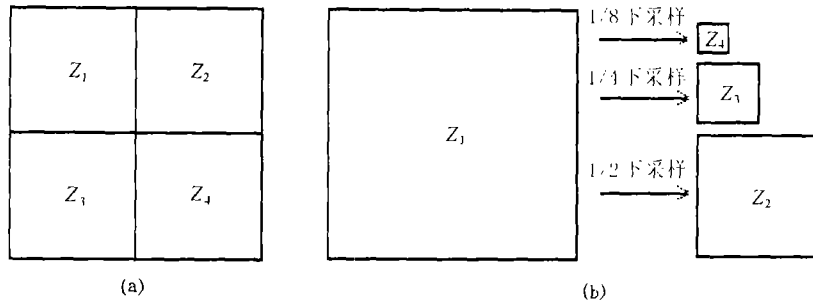


图 4 特征矢量的组织  
(a) 第 1 种特征组织方法图 (b) 第 2 种特征组织方法图

### 3.3 分类器的选择

在纹理分类的过程中,除了确定分类特征外,分类器的选用也是一个关键问题,这里不准备对分类器做过多探讨,而直接使用自组织特征映射神经网络(SOFM)的方法作为分类器。SOFM的学习是一种无监督的聚类方法,与传统的模式聚类方法相比,它所形成的聚类中心能映射到一个曲面或平面上,且保持拓扑结构不变<sup>[9]</sup>。

### 3.4 压缩域的实现方法

上面讨论了纹理分类的 3 个必要步骤,但如果在压缩域操作,还必须考虑一些特殊处理。因为是对 RLE 码流进行操作,故要对图 2(b) 的模板矩阵  $M$  进行 Zigzag 排序,得到与 RLE 码流中 DCT 系数排列顺序一致的矩阵  $B$  :

$$B = \begin{bmatrix} 0 & 1 & 3 & 3 & 2 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 2 & 1 & 1 & 1 \\ 1 & 2 & 2 & 3 & 3 & 3 & 3 & 3 \\ 2 & 1 & 1 & 1 & 1 & 1 & 1 & 2 \\ 2 & 3 & 3 & 3 & 3 & 3 & 3 & 2 \\ 1 & 1 & 1 & 1 & 1 & 2 & 2 & 3 \\ 3 & 3 & 2 & 2 & 2 & 1 & 1 & 2 \\ 2 & 3 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

对于 JPEG 压缩数据,首先进行 Huffman 解码得到 RLE 码流(将 DC 值也按 RLE 格式排列,即 runlength=0, codevalue=DC 值,这样便于程序处理),再直接在 RLE 码流上进行操作处理,处理时可直接跳过 runlength 个零值,只对非零值 codevalue 进行运算,因此运算量比较少,大大节约了处理时间。在特征矢量的组织中,把每个 DCT 块看作一个点值进行处理,求每个  $Z_i$  组的能量  $E_0 \sim E_3$  可分两步,第 1 步先求平均值  $A_0 \sim A_3$ ,第 2 步再求能量  $E_0 \sim E_3$ 。下面的一小段 C++ 程序就是在 RLE 码流中直接求一个  $8 \times 8$  DCT 块中区域能量  $E_0 \sim E_3$  的例子:

```
// 在其它地方定义的数据结构及意义
//typedef struct RLE-Btag {int runlength,codevalue;
//                               RLE-Btag *nextptr;}RLE-B;
//float A[4] 是 DCT 域中四个区域的各自平均值,在调用此函数前已求得
//float E[4] 代表各个区域的能量值,已有初值
void GetEx(float E[4], float A[4], RLE-B *Rleblock)
{int r = 0, k = 0; float c = 0;
```

```

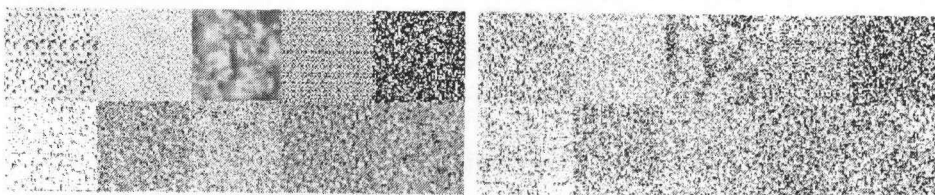
// 注: 熵解码时, 已将 DC 值分给了链表 Rleblock 的第一项: runlength=0; codevalue=DC
值 do
    {k+=Rleblock->runlength;          // 直接跳过值为零的项
    r = B[k/8][k%8];          // 确定非零值所处的区域
    c=(float)Rleblock->codevalue-A[r];
    E[r] += c*c;
    Rleblock=Rleblock->nextptr;      // 处理链表的下一项
    k++;
    }while ((Rleblock!=NULL)&&(k <= 63));
}

```

#### 4 实验结果与结论

根据上面的讨论, 作者对用 JPEG 压缩的纹理图像进行了实验模拟。实验中的纹理图像取自 Adobe 纹理图像库, 共 10 类: Bark, Caviar, Cloud, Denimz, Frzrain, Puzzle, Rustflks, Strands2, Thekhair, Weave3, 分辨率为  $512 \times 512$ 。为了获得足够的训练样本和测试样本, 将原始的  $512 \times 512$  图像以  $1/2$  抽样形成 4 幅  $256 \times 256$  的图像共 40 幅, 将其中的 20 幅作为训练样本, 剩下的 20 幅作为测试样本。对所有的图像进行 JPEG 压缩 (DCT 量化矩阵、质量因子一定, SNR 的平均值为 28.12dB, 恢复图像如图 5(a) 所示), 以便在压缩域直接进行纹理分类。为了验证此分类方法的抗噪能力, 利用 Adobe PaintShop 中的 Random 加噪方法, 给所有的 40 幅图像以 10%, 20%, 30%, 35%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 100% 的力度加噪 (如图 5(b) 所示), 作为测试图像进行测试 (共 440 多幅)。实验中, 采用了 3.2 节中讨论的两种组织特征矢量的方法进行测试, 识别结果如表 2 所示。从表中可以看出, 在 35% 以下的噪音环境中, 两种方法的识别率均为 100%; 噪音在 70% 到 80% 时, 方法 1 的识别率从 90% 陡降为 67.5%, 而方法 2 仍为 90%, 这是因为此时的噪音已严重干扰了原图像的局部特性, 但对其整体特性的影响并不太大; 噪音为 100% 时, 方法 2 的识别率还能保持 82.5%, 但方法 1 已不到 50%。虽然第 1 种方法的总体识别率比不上第 2 种方法, 但是它强调的是局部特性, 在对图像的 AOI(感兴趣区) 分析处理应用中可给予较高权值, 提高分析处理的准确度。

本文提出的基于 DCT 压缩域的纹理图像分类方法是直接在 RLE 压缩域中处理纹理图像的, 对 JPEG 压缩图像只需进行少量的解码 (Huffman 解码), 要处理的数据量也较少, 大大缩短了处理时间; 在 DCT 域中, 本文提出了区域能量具有方向性的特点, 并利用这一特点组织了图像在 DCT 域中的特征矢量, 进行了压缩域纹理图像分类的实验。实验结果表明这种基于 DCT 压缩域的纹理图像分类方法具有较高的正确性和抗噪性。



(a) JPEG 图像

(b) 加了 100% 噪音的纹理图像

图 5 10 类纹理图像

表 2 噪音环境下两种方法对 10 类纹理图像的分类识别率

加噪程度 (%)	<35	40	50	60	70	80	90	95	100
方法 1 识别率 (%)	100	95	95	90	90	67.5	55	47.5	47.5
方法 2 识别率 (%)	100	100	90	92.5	90	90	87.5	87.5	82.5

## 参 考 文 献

- [1] 庄越挺, 潘云鹤等, 基于内容的图像检索综述, 模式识别与人工智能, 1999, 12(2), 170-177.
- [2] 向辉, 石教英, 压缩域多媒体数据处理技术研究, 中国图像图形学报, 1999, 4(7), 539-543.
- [3] B. C. Smith, L. A. Rowe, Algorithms for manipulating compressed images, IEEE Computer Graphics and Applications, 1993, 135(9), 34-42.
- [4] Shih-Fu Chang, Compressed-domain techniques for image/video indexing and manipulation. IEEE ICIP'95 Special Issue on Intelligent Signal Processing, Washington DC, USA, 1995, 314-317.
- [5] Shih-Fu Chang, Some new algorithms for processing images in the transform compressed domain. Proc SPIE Visual Comm. and Image Processing, Bellingham, WA USA, 1995, 445-454.
- [6] 沈兰荪, 魏海, 黄祥林, 压缩域图像处理技术研究, 北京工业大学学报, 2000, 26(3), 24-32.
- [7] B. Chiprasert, K. R. Rao, Discrete cosine transform filtering, Signal Processing, 1990, 19(3), 233-245.
- [8] 徐朝伦, 王晓湘等, 基于子波变换的纹理图像分类, 电子科学学刊, 1999, 21(3), 404-407.
- [9] 胡守仁主编, 神经网络导论, 长沙国防科技大学出版社, 1992, 205-215.

TEXTURE-IMAGE CLASSIFICATION IN  
DCT COMPRESSED-DOMAIN

Huang Xianglin    Shen Lansun

*(Signal and Info. Processing Lab, Beijing Polytechnic University, Beijing 100022, China)*

**Abstract** A novel method of texture-image classification in DCT compressed-domain is proposed. The textural feature of energy-orientation is extracted in DCT transformed-domain and this textural feature can be used to classify texture-images in RLE bit-stream directly. The experimental results demonstrate that texture-images can be classified at higher correct classification rate under random noise.

**Key words** Image processing in compressed-domain, DCT, Texture-image classification, Energy-orientation

黄祥林: 男, 1967 年生, 讲师, 博士生, 研究方向: 智能化信息处理.

沈兰荪: 男, 1938 年生, 教授, 博士生导师, 研究领域: 图像编码与传输, VLSI 实时信号处理, 谱信号检测等.