

# 基于感知度量的分形编码\*

张元亮 郑南宁

(西安交通大学人工智能与机器人研究所, 西安 710049)

**摘要** 传统的分形编码利用最小均方差准则(MMSE)来度量图象子块的变换匹配,未能充分利用人眼的视觉特征。本文提出了基于人眼视觉模型的分形编码方案。算法中,按视觉上的差异对图象子块进行分类;由“视觉最相似”的准则确定最佳匹配的域块,为此,相似匹配转换到DCT域进行,导出了图象块的相似匹配在DCT域的形式;算法进一步通过自适应误差校正来消除解码图象中的方块效应。解码图象的视觉质量有了明显的改善。实验结果证实了算法的有效性。

**关键词** 人眼视觉模型,分形编码,DCT变换

## 1 引言

随着数字图象的广泛应用,涌现出许多新的图象压缩算法,分形图象压缩就是其中一例。Jacquain首先提出基于分块迭代函数系(PIFS)的实用分形编码方案<sup>[1,2]</sup>。认为图象中的信息冗余表现为图象的局部自相似性。为此,将图象划分成值块(Range block)和域块(Range Block,域块构成图象中的虚拟码书),由最小均方差准则(MMSE)在虚拟码书中搜索与值块最相匹配的域块。保存匹配变换而非图象灰度的信息,从而达到了图象压缩的目的。

分形图象压缩从新的方面揭示和利用了图象中的冗余信息,越来越引起人们的重视。但是大量的实验表明,尽管分形编码可以取得较高的压缩比,图象的质量却并不总是令人满意。这是由于(1)传统的分形编码将图象中自相似性假设为值块与域块之间的相似匹配,值块与域块通常为具有规则形状的矩形块,并且其象素点是线性对应的。这样的模型假设有时并不能反映图象中的实际情况。为此人们开始研究基于区域的非规则值块与域块之间的匹配<sup>[3]</sup>;以及值块与域块之间对应点的非线性对应<sup>[4]</sup>。(2)传统的分形编码未能充分利用人眼的视觉特性。例

如,传统的分形编码以最小均方差准则(MMSE)来衡量值块与域块之间的相似性,不符合人眼的视觉特征,人眼对图象信息的处理并不是逐点来进行的,而是抽取空间,频率或色彩的特征进行神经编码。

本文提出了一种新的分形压缩算法。在传统分形编码的框架下,结合人眼的特性进行图象中域块与值块的匹配。根据图象子块视觉上的差异进行分类;通过引入人眼视觉频率响应函数,给出值块与域块“视觉上最相匹配”的准则,推导出了空间域上图象块的压缩变换在DCT域的对应形式。通过基于视觉原则的自适应误差校正来消除方块效应,进一步提高解码图象的质量。实验中,基于视觉感知的分形编码算法用于Lena图象的压缩,取得了较好的视觉效果,证实了算法的有效性。

## 2 人眼视觉系统模型(HVS)

在过去20年里,人们提出了各种各样的视觉模型。但由于模型的复杂性和本质的一些未知因素等条件的制约,这些模型不适合于图象压缩场合<sup>[5,6]</sup>。大量的研究发现,当用于图象质量的评判等场合时,一些简单视觉模型仍能达到优于MMSE的效果。文

\* 国家自然科学基金和杰出青年科学基金资助项目

收稿日期:1997-09-18;收到修改稿日期:1997-12-16

献[7]中提出的一种视觉系统的频率响应为:

$$H(\omega) = (a + b\omega)\exp(-c\omega) \quad (1)$$

其中  $\omega$  为视角正对的径向频率,单位为周/度,  $a, b, c$  为决定 HVS 曲线形状的常数。当  $\omega_{\max} = 3$  周/度时, HVS 的曲线形状为:

$$H(\omega) = (0.31 + 0.69\omega)\exp(-0.29\omega) \quad (2)$$

在图象编码中应用 DCT 变换时,相当于对原始图象进行对称扩展,但是人眼却观测不到这一并不存在的场景变换。文献[7]中提出矫正函数:

$$|A(\omega)| =$$

$$\left[ \frac{1}{4} + \frac{1}{\pi^2} \left( \ln \left( \frac{2\pi\omega}{\alpha} + \left( \frac{4\pi^2\omega^2}{\alpha^2} + 1 \right)^{1/2} \right) \right)^2 \right]^{1/2} \quad (3)$$

其中  $\alpha = 11.636$  度,这样视觉系统的响应函数可以写成:

$$\tilde{H}(\omega) = |A(\omega)|H(\omega) \quad (4)$$

采用如下的公式<sup>[8]</sup>,将二维的 DCT 系数  $(u, v)$  转换为径向频率  $\omega$

$$\omega(\text{cycle/degree}) = \omega_d(\text{cycle/pixel}) \times \omega_s(\text{pixels/degree}) \quad (5)$$

其中  $\omega_d = \sqrt{u^2 + v^2} / 2N, u, v = 0, 1, \dots, N-1, \omega_s$  是依赖于观测距离的采样函数,  $N$  是 DCT 块的大小,对图象高度为 288 个象素点 CIF 图象,当观测距离位于图象高度的 4 倍处时,  $\omega_s = 20$  象素点/度,在我们的实验中,取  $\omega_s = 32$ 。

### 3 基于视觉模型的分形编码

基于视觉感知的分形编码算法的结构如图 1。

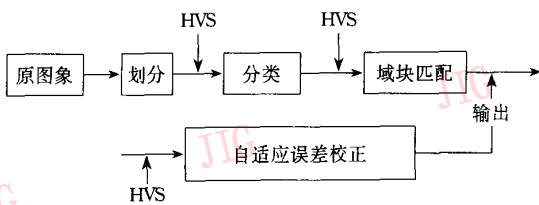


图1 基于视觉感知的分形编码

为了降低算法的复杂度和提高解码图象的质量,分形编码中对值块和域块进行分类,只对属于同一类的图象块进行匹配。传统的分类方法采用 Ger-sho<sup>[9]</sup>在分类矢量量化(CVQ)中的方法。但空间域的这种分类方法只对  $4 \times 4$  和  $5 \times 5$  的小块有效,不适合  $8 \times 8$  或  $16 \times 16$  等较大的图象子块。在基于视觉模型的分形编码中,根据视觉上的差异对图象子块进行分类。为此,定义视觉交流总能量:

$$T = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \tilde{H}(u, v) |F(u, v)| - \tilde{H}(0, 0) |F(0, 0)| \quad (6)$$

视觉低频交流能量:

$$L = \sum_{u=0}^{N/2-1} \sum_{v=0}^{N/2-1-u} \tilde{H}(u, v) |F(u, v)| - \tilde{H}(0, 0) |F(0, 0)| \quad (7)$$

低频交流能量占总能量之比:

$$r = L/T \quad (8)$$

相对平坦的子块,其 DCT 谱的交流分量接近零。而对于复杂的子块,其 DCT 谱的交流分量则占有相当的成分。可以认为,视觉交流总能量  $T$  反应了图象子块视觉复杂性,根据  $T$  的取值可以对图象块作初步的分类。对于非平坦的子块,存在或边缘或纹理的特征。单一边缘的子块,其 DCT 谱的交流能量集中在低频区域<sup>[10]</sup>,能量比  $r$  进一步刻画出子块的单一边缘或复杂边缘(纹理)特征。

实验中,对 Lena 图象的分类结果如图 3 所示。根据 Lena 图象子块的统计特性选取阈值为  $Thresh1 = 13, Thresh2 = 0.68$ ,其中,黑色区域为平坦块,灰色区域为边缘块,白色区域为复杂块。可见,基于视觉感知的分类基本上正确反应图象中不同类子块的分布。

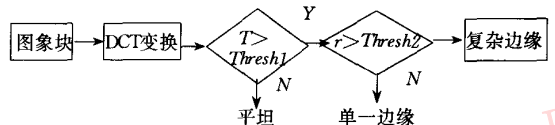
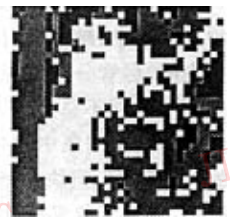


图2 基于视觉感知的图象子块分类



(a) Lena 图象



$Thresh1 = 13$   
 $Thresh2 = 0.68$   
(b) Lena 图象分类图

图3 Lena 图象的分类图

在基于视觉感知的分形编码中,对平坦块,只保存其灰度的均值信息进行量化编码。对边缘和纹理类中的每一值块  $f_R(x, y)$ ,在同类的虚拟码书中搜

索域块  $f_D(x, y)$ , 满足或近似满足匹配关系:

$$f_R(x, y) = P \cdot S \cdot T(f_D(x, y)) \quad (9)$$

式(9)中,  $T$  为域块的旋转与对称变换,  $S$  为域块的收缩变换,  $P$  为域块到值块的仿射变换  $P(f_D(x, y)) = af_D(x, y) + b$ , 其中  $a$  需满足使仿射变换最终收敛<sup>[11]</sup>的条件。

传统的分形编码对给定的值块, 按 MMSE 准

$$F_R(u, v) = aF_D(u, v) + \frac{1}{16}c(u)c(v) \sum_{x=0}^7 \sum_{y=0}^7 bc \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} = \begin{cases} aF_D(u, v) + 2b & (u, v) = (0, 0) \\ aF_D(u, v) & \text{otherwise} \end{cases}$$

对于给定的值块  $f_R(x, y)$ , 定义与其“视觉意义下最匹配的子块”为满足:

$$\min \sum_{\substack{u=0 \\ (u,v) \neq (0,0)}}^7 \sum_{v=0}^7 H(u, v)^2 (F_R(u, v) - aF_D(u, v))^2 + H(0, 0)^2 (F_R(0, 0) - aF_D(0, 0) - 2b)^2 \quad (10)$$

的域块。由最小二乘法, 达到极小值的  $a, b$  满足:

$$a = \frac{\sum_{\substack{u=0 \\ (u,v) \neq (0,0)}}^7 \sum_{v=0}^7 H(u, v)^2 F_R(u, v) F_D(u, v)}{\sum_{\substack{u=0 \\ (u,v) \neq (0,0)}}^7 \sum_{v=0}^7 H(u, v)^2 F_D(u, v)^2} \quad b = \frac{1}{2} (F_R(0, 0) - aF_D(0, 0)) \quad (11)$$

从式(11)可以看出,  $a$  的选取与图象子块的直流量无关, 可以抽取图象子块中的直流量单独进行编码。这等价于图象子块的正交化过程<sup>[12]</sup>。式(9)中,  $T$  为域块的旋转与对称变换, 文献[1, 2]中给出了域块空间域中的 8 种旋转与对称变换, 类似于

$$(S \cdot F_D)(u, v) = \frac{1}{16}c(u)c(v) \sum_{x=0}^7 \sum_{y=0}^7 \frac{1}{4} (f_D(2x, 2y) + f_D(2x+1, 2y) + f_D(2x, 2y+1) + f_D(2x+1, 2y+1)) \times \cos \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \quad \text{其中 } u, v = 0, 1, \dots, 7 \quad (12)$$

而

$$F_D(u, v) = \frac{1}{16}c(u)c(v) \sum_{x=0}^7 \sum_{y=0}^7 \frac{1}{4} (f_D(2x, 2y) \cos \frac{(4x+1)u\pi}{32} \cos \frac{(4y+1)v\pi}{32} + f_D(2x+1, 2y) \cos \frac{(4x+3)u\pi}{32} \cos \frac{(4y+1)v\pi}{32} + f_D(2x, 2y+1) \cos \frac{(4x+1)u\pi}{32} \cos \frac{(4y+3)v\pi}{32} + f_D(2x+1, 2y+1) \cos \frac{(4x+3)u\pi}{32} \cos \frac{(4y+3)v\pi}{32}) \quad (13)$$

展开式(13)中各项, 可得:

$$F_D(u, v) = (S \cdot F_D)(u, v) \cos \frac{u\pi}{32} \cos \frac{v\pi}{32} + \frac{1}{64}c(u)c(v) \sum_{x=0}^7 \sum_{y=0}^7 (f(2x, 2y) - f(2x, 2y+1) + f(2x+1, 2y) - f(2x+1, 2y+1)) \cos \frac{(2x+1)u\pi}{16} \sin \frac{(2y+1)v\pi}{16} \cos \frac{u\pi}{32} \sin \frac{v\pi}{32} + \frac{1}{64}c(u)c(v) \sum_{x=0}^7 \sum_{y=0}^7 (f(2x, 2y) + f(2x, 2y+1) - f(2x+1, 2y) - f(2x+1, 2y+1)) \sin \frac{(2x+1)u\pi}{16} \cos \frac{(2y+1)v\pi}{16} \sin \frac{u\pi}{32} \cos \frac{v\pi}{32} + \frac{1}{64}c(u)c(v) \sum_{x=0}^7 \sum_{y=0}^7 (f(2x, 2y) - f(2x, 2y+1) - f(2x+1, 2y) +$$

则, 由最小二乘法计算最佳匹配的子块。MMSE 只是计算对应象素点之间差异的总和, 未能充分考虑人眼的视觉特征。从上面引入的简单的视觉模型可以看出, 人眼对不同的频率具有不同的视觉敏感性, 因而值块与域块之间视觉意义下的最佳匹配转换到 DCT 域进行。仿射变换  $f_R(x, y) = P(f_D(x, y)) = af_D(x, y) + b$  对应在 DCT 域形式为:

文献[13], 可以得到 8 种变换对应在 DCT 域的形式。

式(12)中,  $S$  为域块到值块的收缩变换。在分形编码中, 域块需大于值块, 经收缩变换后才与值块对应, 收缩变换后图象子块的 DCT 系数为:

$$f(2x + 1, 2y + 1)) \sin \frac{(2x + 1)u\pi}{16} \sin \frac{(2y + 1)v\pi}{16} \sin \frac{u\pi}{32} \sin \frac{v\pi}{32} \quad u, v = 0, 1, \dots, 7 \quad (14)$$

由于  $|\cos \frac{k\pi}{32}| > |\sin \frac{k\pi}{32}|$  对于  $0 < k < 8$ , 并且  $|f(2x, 2y) + f(2x + 1, 2y) + f(2x, 2y + 1) + f(2x + 1, 2y + 1)| \gg |f(2x, 2y) - f(2x + 1, 2y) + f(2x, 2y + 1) - f(2x + 1, 2y + 1)|$  其余各项类推。因而可以认为:

$$F_D(u, v) \approx (S \cdot F_D)(u, v) \cos \frac{u\pi}{32} \cos \frac{v\pi}{32} \quad u, v = 0, 1, \dots, 7 \quad (15)$$

值块到域块的收缩变换在 DCT 域的对应关系如图 4。

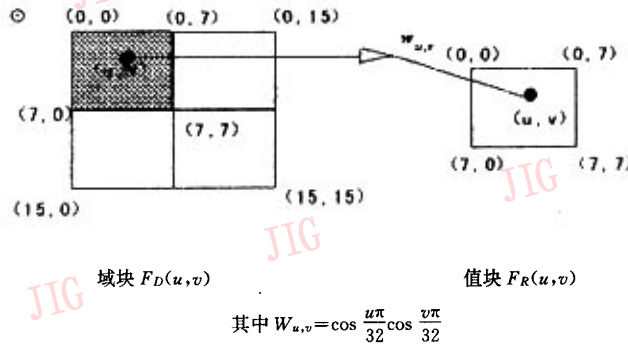


图 4 DCT 域的收缩变换

### 4 自适应误差校正

类似于所有的块状编码压缩算法, 在达到较高的压缩比时, 子块边缘可能出现方块效应。在本文的算法的解码图象中, 每个子块的失真度量反映了此图象块的重构质量, 如果两个或多个相邻的图象子块之间的失真梯度较大, 人眼将会观察到块间不连续性的存在, 导致了图象中方块效应和虚假边缘的出现。消除方块效应的途径之一是降低块间的失真梯度。定义图象子块  $(m, n)$  的视觉失真误差为  $E(m, n) = \sum_{u=0}^7 \sum_{v=0}^7 \tilde{H}(u, v)^2 [F(u, v) - \tilde{F}(u, v)]^2$ 。其中,  $F(u, v)$  和  $\tilde{F}(u, v)$  分别是原始图象和解码图象的 DCT 分量,  $\tilde{H}(u, v)$  为如上所述的 DCT 域的人眼频率响应函数。定义块间失真比为  $G(m, n) = \log_{10}[E(m, n)/E(m, n-1)]$ 。其中,  $E(m, n)$  和  $E(m, n-1)$  分别是图象当前子块和其前一子块的视觉失真度量, 正的  $G(m, n)$  表明当前子块的失真度量大于其前一子块; 反之, 则相反。 $|G(m, n)|$  反映了相邻的图象子块的失真度量之间的相对关系。选取一经验阈值, 如果超过这一阈值, 则需要对子块进行误差均衡; 否则, 认为子块的失真是可以接受的。影响  $|G(m, n)|$  阈值选取的一个因素是需要进行误差均衡的图象子块的数目。为达到满意的图象压缩比,

必须控制进行失真均衡的图象子块数目。对于误差图象子块, 采用类似于 JPEG 中用到的编码策略, 误差图象子块的 DCT 系数量化后按 zigzag 的顺序排列。对于其中的非零系数, 由幅值查找表进行编码, DCT 系数的查找表地址经过游程编码后, 再进行 Huffman 编码进一步提高压缩比。

相对于编码过程, 解码过程比较简单。类似于经典的分形编码, 解码过程仍然在空间域中进行。对于编码的误差图象子块, Huffman 及游程反编码后, 经过反量化即可恢复。

### 5 实验结果和结论

基于视觉感知的分形编码中, 域块由水平和垂直方向各 4 个像素点的平移而得到。压缩图象变换的系数  $a$  (如式(11)所示) 在  $0.2 \sim 0.9$  之间取值, 系数  $b$  以及域块位置的量化位数分别为 6 位和 5 位。按文献[14]中提出的快速算法进行 DCT 变换, 预先计算并保存值块和域块的 DCT 系数值, 匹配过程中只需查表运算。假设图象中值块的大小为  $N = 8 \times 8$ , 值块的数目为  $M_R = 256 \times 256 / N$ , 域块的数目为  $M_D$ 。与传统的分形编码相比, 值块与域块一次匹配的计算复杂度如表 1 所示。传统的分形编码按 Fisher 给出<sup>[11]</sup>的形式计算。

表 1 一次匹配的算法复杂度比较

|    | 快速 DCT                   | 本文方法     | 分类分形编码    |
|----|--------------------------|----------|-----------|
| 加法 | $1.5N(\log_2 N - 1) + 2$ | $3N + 1$ | $7N + 5$  |
| 乘法 | $N \log_2 N - 1.5N + 4$  | $N + 3$  | $3N + 13$ |

按照上面的参数设置进行实验,则本文算法编码过程的总计算量分别为:

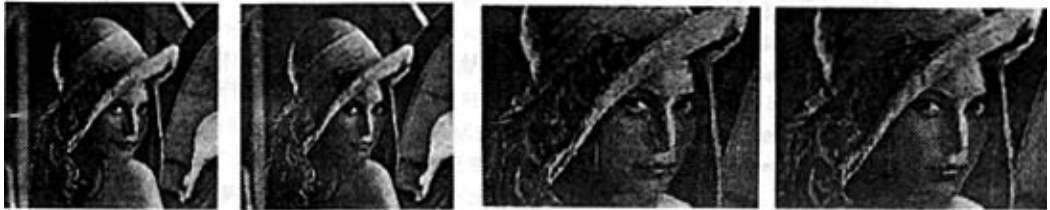
$$\text{加法: } (3N + 1)M_n M_b + (1.5N(\log_2 N - 1) + 2)$$

$$(M_D + M_R)$$

而经典的分形编码的计算量为:

$$\text{加法: } (7N + 5)M_D M_R, \text{乘法: } (3N + 13)M_D M_R.$$

因而,本文算法节约计算量一半以上,以空间存储上的代价换取了时间上的收益。对于较大的图象可以将其划分成子图象,独立地进行分形编码。分形编码后的重构图象(失真校正之前)如图 5 所示,解



(a) 经典分形编码的重构图象 (b) 本文算法的重构图象 (c):(a)图的局部放大 (d):(b)图的局部放大 (未作误差校正)

图 5 分形编码的解码图象

解码图象的峰值信噪比 PSNR 分别为:经典方法 28.16,本文算法 27.85。可以看出,与传统的分形编码相比,本文算法(失真校正之前)重构图象的视觉质量有了一定的改善。说明了 MSE 与图象的主观质量有时是不完全一致的。实验中  $G(m, n)$  取为 0.25,进行失真均衡的图象块的数目为 283。通过计

算  $G(m, n)$  可以自适应地对子块进行失真均衡。自适应失真均衡阈值的大小决定了需进行误差均衡的图象块的数目。分形编码中结合人眼视觉模型,使解码图象的主观质量有了一定的改善。自适应失真校正进一步地提高了重构图象的质量(图 6)。



(a) 进行失真均衡的误差 灰度上移 128 (b) 本文算法的解码图 压缩比为 0.47bpp PSNR=30.2 (c) JPEG 压缩图象 压缩比为 0.47bpp PSNR=29.1

图 6 压缩效果比较

## 参考文献

- 1 Jacquin A E. Image coding based on a fractal theory of iterated contractive image transformations. *IEEE Trans. IP*, 1992, 1(1): 18~30.
- 2 Jacquin A E. Fractal image coding a review. *Proceedings of IEEE*, 1993, 81, 1451~1465.
- 3 Thomas L, Deravi F. Region-based fractal image compression using heuristic search. *IEEE Trans. IP*, 1995, 4: 832~838.
- 4 Popescu D C, Dimca A, Yan H. A Nonlinear Model for fractal image coding. *IEEE Trans. on IP*, 1997, 6, 373~382.
- 5 Overington I. Toward a complete model of photopic visual threshold performance. *Opt. Eng.*, 1974, 21: 2~13.
- 6 Schitzler A. Effects of spatial frequency filtering on the performance of the composite photographic-human visual system. *Photo. Sci. Eng.* 1997, 21: 209~215.
- 7 Nill N B. A visual model weighted cosine transform for image compression and quality assessment. *IEEE Trans. COM*, 1985, 33: 551~557.
- 8 Tan S H, Ngan K N. Classified perceptual coding with adaptive quantization. *IEEE Trans. on Circuits and System for Video Technology*, 1996, 6: 375~388.
- 9 Ramaurthi B, Gersho A. Classified vector quantization of images. *IEEE Trans. COM*, 1986, 34.
- 10 Kim D S, Lee S U. Image vector quantizer based on a classification in the DCT domain. *IEEE Trans. COM*, 1986, 34.
- 11 Fisher Y. *Fractal image coding-theory and application*. Springer-Verlag, 1994.
- 12 Oien G E, Lepsoy S. Fractal-based image coding with fast decoder convergence. *Signal Processing*, 1994, 40: 105~117.
- 13 Bracewell R N, K. -Y, et al. Affine theory for two dimensional fourier transform. *Electron. Letter*, 1993, 29: 304.
- 14 Chen W H, Smith C H, Fralick S C. A fast computational algorithm for the discrete cosine transform. *IEEE Trans. COM*, 1997, 25: 1004~1009.



张元亮 男, 1970年3月生, 1991年毕业于北京大学数学系, 现为西安交通大学人工智能与机器人研究所读博士生, 主要从事图象编码, 模式识别方面的研究, 已发表论文多篇。



郑南宁 工学博士, 教授, 博士生导师。IEEE高级会员。1952年生于南京, 1975年10月毕业于西交大工业自动化专业, 1981年获本校自控理论与应用专业硕士, 1985年3月获日本庆应大学工学博士学位, 先后发表论文百余篇, 编著有《数字信号处理》, 计算机视觉与模式识别, 现主要从事机器视觉与模式识别, 神经网络, 并行处理的研究。

## A Fractal Image Coding Based on Visual perception

Zhang Yuanliang      Zheng Nanning

(Institute of Artificial Intelligence and Robotics Xi'an Jiaotong University Xi'an 710049)

**Abstract** In classical fractal image coding, the self-similarities between image blocks are measured by MMSE. No human visual properties are considered. In this paper, we propose a novel fractal block coding based on human visual system. The original image blocks are classified by their visual differences; For each range block, the domain block which is "the most similar in vision" is searched and the matching of blocks is completed in the DCT domain; Adaptive distortion equalization is used to reduce the blocking effect in the decoded image. The performance of our method has a visible improvement of subject quality depending on visual perception.

**keywords** Human visual model, Fractal block coding, DCT transform