

基因表达式编程在分形图像压缩编码中的应用

李文敬, 刘之家, 廖伟志

(广西师范学院 计算机与信息工程学院, 广西 南宁 530001)

摘 要: 针对目前存在图像压缩率不高、全局搜索或遗传算法寻找最优分形图像压缩编码速度慢的不足, 将基因表达式编程应用于分形图像压缩编码, 提高求解分形图像压缩编码速度和压缩比。首先从理论上对二值图像压缩编码的求解过程和基因表达式编程在压缩编码中的作用机理进行分析; 然后, 研究分形图像压缩编码的基因和染色体的表示方法, 适应度函数设计以及选择、变异、插串、基因变换、基因重组等基因遗传进化操作过程。提出基因表达式编程的分形图像压缩算法, 求解分形图像压缩编码的最优解。实验结果表明, 基因表达式编程应用于分形图像压缩编码, 具有较强的全局寻优能力, 搜索最优解的速度比遗传算法快约 2 倍, 图像压缩率高。

关键词: 基因表达式编程; 分形图像; 迭代函数系统; 遗传进化; 压缩编码

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2011)08-0067-03

Gene Expression Programming Applied to Fractals Image Compression Encoding

LI Wen-jing, LIU Zhi-jia, LIAO Wei-zhi

(College of Computer Science and Information Engineering,
Guangxi Teachers Education University, Nanning 530001, China)

Abstract: Based on the existing image compression rate was not high and global search or genetic algorithm to find the optimal fractal image compression coding deficiencies slow at present. Gene expression programming applied to fractals image compression encoding, so that improve the speed and compression ratio of fractal image compression. First, binary fractal image compression process of solving IFS was analyzed; the gene and chromosome structure and coding of fractal image compression IFS of gene expression programming were given. According to the similarity of fractal images, compression and shrinkage characteristics, and the fitness function of gene expression programming and genetic evolution operating of select, mutating, insert string, gene transformation, gene recombination were given. The fractals image compression algorithm based on genetic expression programming was put forward. The experimental results show that the algorithm has a fairly strong global search optimization capacity, convergence rate of about twice times faster than based on genetic algorithm and high compression ratio.

Key words: gene expression programming; fractals image; iterated function system; genetic evolution; compression encoding

1 引言

Barnsley 对几幅特定图像进行分形压缩编码, 获得了 10 000 : 1 的压缩比, 显示出分形技术在图像压缩编码方面的巨大潜力。在二值图像分形压缩

研究方面, Dipankar Dasgupta^[1] 与刘冠蓉教授^[2] 分别提出了基于遗传算法的二值图像压缩算法。他们通过遗传算法在全局最优解搜索方面的优点, 提高分形压缩的编码速度。但是, 遗传算法的个体存活率不高, 检查个体存活浪费大量的时间, IFS 的收敛速

仿射变换的个数,其他参数及公式的意义参见文献[1].

交叉相似度量公式如下:

$$s(B, \bigcup_{i=1}^n \omega_i(B)) = \left| \frac{B \cap (\bigcup_{i=1}^n \omega_i(B))}{B \cup (\bigcup_{i=1}^n \omega_i(B))} \right| \quad (7)$$

压缩函数 $c(\lambda)$ 定义:

$$c(\lambda) = (1 - \lambda^{10})e^{\frac{\lambda^2}{4\eta^2}} \quad (8)$$

仿射变换个数 $L(\xi)$ 函数定义:

$$L(\xi) = e^{-\frac{\mu^2}{4\xi^2}} \quad (9)$$

3.5 GEP 的遗传操作

GEP 的遗传基本算子包括选择、变异、倒串、插串、根插串、基因变换、单点重组、两点重组和基因重组等 9 种遗传进化操作. 遗传操作方法请参看文献[5].

4 基于 GEP 分形图像压缩算法

根据第 3 节中的分析与讨论,提出了基于 GEP 分形图像压缩算法,算法描述如下:

输入:种群大小、基因头部尾部长度、基因个数、最大迭代数、终止迭代适应度值、变异率、插串率、重组率等初始参数;

输出:最优的分形迭代函数系统 IFS.

Step1:种群初始化;

Step2:读原图像数据;

Step3:适应度值的计算;

Step4: While(fitness <= Maxf Or gen < Maxg)

{ Step5:选择操作; Step6:变异操作;

Step7:倒串操作; Step8:插串操作;

Step9:根插串操作; Step10:基因变换;

Step11:单点重组; Step12:两点重组;

Step13:基因重组; Step14:个体的适应度值的计

算; }

输出最优编码 IFS;

5 实验结果与分析

在 CPU 为 P4, 2.8 GHz, 内存 512 MB 的 PC 机上,用 Microsoft Visual C++ 2005 编程实现. 原始二值图像是 128×128 的 C 曲线,如图 1 所示,图像大小为 11.6 kB.

基于遗传算法和 GEP 分形图像压缩算法实验的主要参数设置如表 1 所示.

基于遗传算法和 GEP 的分形图像压缩算法实验得到的最优 IFS 解 ω_1 和 ω_2 , 分别如表 2、表 3

所示.

表 1 GA 算法和 GEP 算法主要参数经验值设置

GA 参数描述	值	GEP 参数描述	值
Pop_size 初始种群	50	Pop_size 初始种群	50
Gap 下一代个体	5	Gap 下一代个体	5
p_c IFS 杂交概率	0.5	变异率	0.044
p_m IFS 变异概率	0.9	倒串,插串,根插串,变换	0.1
ξ 压缩仿射变换个数	4	单点重组	0.4
λ 收缩因子	0.5	两点重组	0.2
Generations 最大代数	10^5	基因重组	0.1

表 2 C 曲线 GA 算法最优 IFS 解 ω_1 和 ω_2

a	b	c	d	e	f
0.500 124	-0.484 755	0.500 014	0.499 214	0.000 154	7-0.000 253
0.474 554	0.511 477	-0.425 844	0.536 544	0.447 854	1 0.414 788 8

表 3 C 曲线 GEP 算法最优 IFS 解 ω_1 和 ω_2

a	b	c	d	e	f
0.491 241	-0.490 174	0.500 626	0.510 234	0.000 110	2-0.000 021
0.464 512	0.500 477	-0.465 269	0.526 545	0.457 154	0.474 718 3

实验原图像见图 1,基于遗传算法(GA)重构图像见图 2,本文算法解码重构图像见图 3.

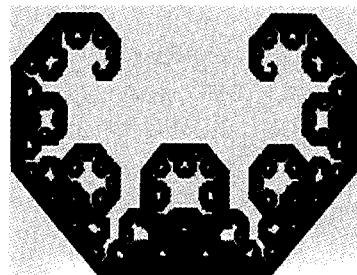


图 1 C 曲线原图像

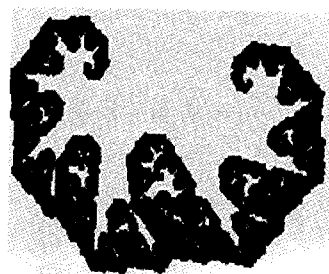


图 2 GA 重构图像

实验结果表明,求解二值图像 C 曲线的分形压缩编码,基于 GA 算法经过遗传进化 680 代,本文算法经过进化 350 代,得到最好的基因解. 所以, GEP 的搜索速度及进化收敛速度比 GA 算法快约 2 倍, 该算法有一定的优势. 从表 2、表 3 得出, 该 IFS 系

(下转第 73 页)

网格边长 $d=0.5$, 不确定邻域 $\epsilon=0.25$, 平均密度门限 $\text{density}=0.8$. 随着轨迹数量增加, 兴趣区数量增加, 如图5所示.

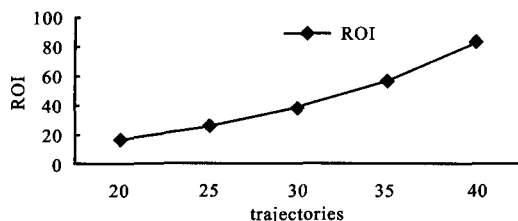


图5 兴趣区域个数与轨迹条数

5 结束语

本文针对轨迹数据的不确定性, 引入模糊集理论, 提出不确定轨迹数据模式挖掘方法. 方法使用网格划分对象运动平面; 针对每个不确定点用隶属度度量该点属于单元格的程度; 统计出每个单元格隶属度的和作为该单元格的密度; 扩展密度单元格得到兴趣区域; 在兴趣区域的基础上挖掘轨迹模式. 最后使用“GeoLife”项目收集的数据进行试验, 展示了本文提出的方法的效果.

参考文献:

[1] Fosca Giannotti, Mirco Nanni, Dino Pedreschi, et al.

Trajectory pattern mining [C]// Proc. KDD, 2007. USA; California, 2007.

[2] Lee Jae Gil, Han Jiawei, Whang Kyu Young. Trajectory clustering: a partition-and-group framework [C]// Proc. ACM SIGMOD Int. Conf. China; Beijing, 2007.

[3] Marcin Gorawski, Pawel Jureczek. Regions of interest in trajectory data warehouse [C]// ACIIDS 2010. Berlin; Springer Verlag, 2010.

[4] Jiong Yang, Meng Hu. TrajPattern: mining sequential patterns from imprecise trajectories of mobile objects [C]// EDBT 2006. Germany; Munich, 2006.

[5] Trajcevski G, Wolfson O, Hinrichs K, et al. Managing uncertainty in moving objects databases [J]. ACM TODS. 2004, 29(3): 463-507.

[6] Nikos Pelekis, Ioannis Kopanakis, Evangelos E Kotsifakos, et al. Clustering uncertain trajectories [M]. Berlin; Springer-Verlag, 2010.

作者简介:

李帆 男, (1986-), 硕士研究生. 研究方向为知识发现与数据挖掘.

夏士雄 男, (1961-), 博士, 教授, 博士生导师. 研究方向为智能信息处理.

张磊 男, (1977-), 博士, 副教授. 研究方向为数据挖掘.

(上接第69页)

统由2个压缩仿射变换 ω_1, ω_2 构成, 即原图像大小为 11.6 kB, 但求解出分形压缩最优解 IFS 只需要存储 $2 \times 6 = 12$ 个实数, 占用 48 B 的存储空间, 压缩率约为 4%.

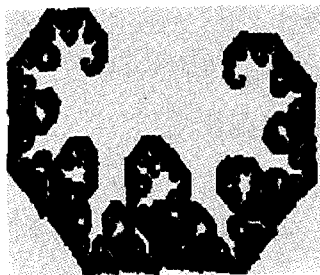


图3 GEP 重构图像

6 结束语

从基因表达式编程具有快速逼近和快速搜索能力出发, 构建 GEP 的分形二值图像压缩编码算法. 通过 GEP 的快速搜索和进化收敛能力求解二值图像最优解, 进化收敛速度比基于 GA 算法快. 而基于 GEP 的彩色图像分形图像压缩编码算法的研究是

我们下一步的研究工作.

参考文献:

[1] Dipankar Dasgupta, German Hernandez, Fernando Niño. An evolutionary algorithm for fractal coding of binary images [J]. IEEE Transactions on Evolutionary Computation, 2000, 4 (2): 172-181.

[2] 刘冠蓉, 郭京蕾, 何华. 基于遗传算法的二值图像压缩 [J]. 武汉理工大学学报: 信息与管理工程版, 2001, 23 (4): 9-13.

[3] Candida Ferreira. Gene expression programming: a new adaptive algorithm for solving problems [J]. Complex Systems, 2001, 13(2): 87-129.

[4] 安建成, 靳海娟. 自适应策略在实数编码遗传算法中的应用研究 [J]. 微电子学与计算机, 2011, 28(4): 140-142.

[5] 元昌安, 彭昱忠, 覃晓, 等. 基因表达式编程算法原理与应用 [M]. 北京: 科学出版社, 2010: 13-95.

作者简介:

李文敬 男, (1964-), 教授. 研究方向为演化算法、图像处理.