

基于基因编码算法的图像轮廓特征信息提取研究

杜广朝¹, 刘 晓², 邵明省²

(1. 黄河水利职业技术学院, 河南 开封 475004; 2. 鹤壁职业技术学院, 河南 鹤壁 458030)

摘 要:针对图像轮廓特征信息提取的特点,提出基因编码算法。首先把像素映射为基因个体表达式,用基因编码值区分为图像轮廓特征,接着对像素的每个基因所编码属性筛选进行特征提取,最后给出了图像轮廓特征信息提取的算法过程。实验仿真得出基因编码提取的图像轮廓边缘很清晰,抗噪声能力强,时效性快,边缘点保持指数最接近1。

关键词:基因编码;分类;轮廓

中图分类号:TP391.4 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2013.07.022

Image contour feature information extraction based on genes encoding algorithm

DU Guang-chao¹, LIU Xiao², SHAO Ming-sheng²

(1. Yellow River Conservancy Technical Institute, Kaifeng 475004, China;

2. Hebi College of Vocation and Technology, Hebi 458030, China)

Abstract: Aiming at the characteristic of the image contour information extraction, the gene coding algorithm is proposed. Firstly, the pixels are mapped to individual gene expression, image contour features are distinguished by using gene encoding values, then each gene encoding feature of the pixel is selected, and the feature is extracted, finally the image contour feature information extraction algorithm is given. The experimental simulation results show that the extracted image contour with gene encoding is very clear, and the algorithm has strong anti-noise ability and quick response. The edge retention index is close to 1.

Key words: gene code; classification; contour

1 前言

图像轮廓特征依赖于图像内容,轮廓特征提取了图像中视觉特征信息,以便对目标区域做进一步处理,如区域填充、图像识别、轮廓长度、面积、重心计算等。

目前图像轮廓特征信息提取研究的方法有: Fourier 变换法、窗口 Fourier 变换(Gabor)、小波变换法、最小二乘法、边界方向直方图法、几何形变模型等。这些方法不能准确对轮廓定位,轮廓信息被弱化,使得后续处理变得较为困难,适用范围也有局限性,会产生很大的误差^[1-3]。图像轮廓特征信息提取目前使用人工智能算法较多,比如

粒子群算法、形态学理论,该类算法参数易设,但是在处理的后期由于数据坍塌,其结果并非是目标函数的最优解^[4]。

本文提出一种基于基因编程算法的图像轮廓特征信息提取,用基因编程算法对像素映射为基因个体表达式,当基因是被表达时则像素点的基因编码值区分为图像轮廓特征,根据每个基因所编码的属性各不相同,通过此对图像的相关属性进行筛选进

基金项目: 河南省教育厅科学技术研究重点项目(No. 12BS10016)资助;鹤壁职业技术学校技术课题(HZY-2012-52)。

作者简介: 杜广朝(1966-),男,硕士,副教授,从事高校电气技术专业教学与研究工作。E-mail: dgchao@foxmail.com

收稿日期: 2012-11-23; **修订日期:** 2012-12-11

行特征提取。实验仿真得出基因编码提取的图像轮廓边缘很清晰,抗噪声能力强,时效性快,边缘点保持指数接近1。

2 基因编码算法描述

2.1 基因表达式编码的基本概念

基因用长度固定的符号串表示,由头部和尾部组成,头部元素包含任何函数符号和终结符号,尾部元素只含终结符号,保证了基因表达结构组织得以保持为合法的计算机程序,基因编码方式的特点可使新个体在结构上延续^[5-6]。

头部长度和尾部长度间的关系为:

$$t = h(n_{\max} - 1) + 1$$

其中, $n_{\max}$ 为函数集中函数的最大参数数目。

在基因编码中的首位位置是基因的起始码,但是结束码子却未必是基因的末位,即基因串的的最后一些位区有可能是非编码区或基因内区。基因的编码表现形式有基因型表达式和表现型 k 表达式树,每个基因对应一个 k 表达式和表达式树, k 表达式和表达式树两者之间可以相互转换^[7-8]。以 $\sqrt{x^2 + y^2}$ 来说明,表达式树从上到下、从左至右的层次遍历,可得到基因型 k 表达式;反之,基因型表达式的逆运算过程可得到表达式树。

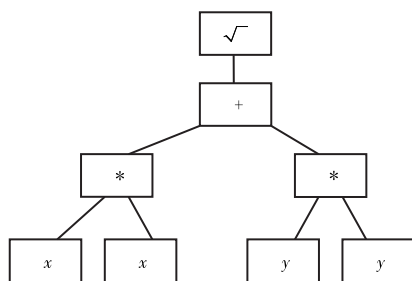


图1 解码生成表达式树

图1中表达式树对应的基因型表达式过程为:

1	2	3	4	5	6	7	8
√	+	*	*	x	x	y	y

基因表达式树里的每1个个体都是1个数学表达式,且都有1个适应度值,根据适应度值对个体进行演化计算,再人为规定1个或几个结果区间作为像素的分类标准。当基因个体表达式的计算值落到某一个区间时,就将其划分为相应的那一类。基因 k 表达式生成表达式树过程为:基因由函数集 $F = \{\sqrt{\quad}, +, -, *, /\}$ 终点 $\{x, y\}$ 构成所以 $n_{\max} = 2$, 假设 $h = 8$, 则 $t = 8 * (2 - 1) + 1 = 9$, 则基因总的长度为 $8 + 9 = 17$, 这样的基因(基因尾部用黑粗体表示)为:

0	1	2	3	4	5	6	7	0	1	2	3	4	5	6	7	0
*	y	+	x	-	x	√	x	y	+	x	y	x	x	x	y	x

则编码生成的表达式树为:

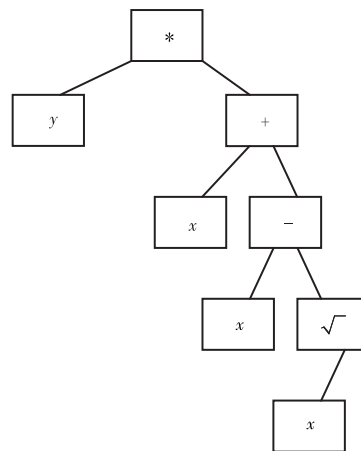


图2 编码生成表达式树结构

在图2的表达式树结构中读取为: $y * [x + (x - \sqrt{x})]$ 。

2.2 基于基因编码算的图像轮廓特征信息分类

图像 $G(V, F)$ 像素由集合 $V = (x_1, x_2, \dots, x_n)$ 和序列 $F = (F_1, F_2, \dots, F_n)$ 组成, 其中 $F_i = \{f_1^{(i)}, f_2^{(i)}, \dots, f_l^{(i)}\}$ 是每个目标基因 x_i 的像素点 i 的编码集合, 集合中 $f_{l(i)}^{(i)}$, 被选中的概率为 $C_{j(i)}^{(i)}$ 。 $x_i \in \{0, 1\}$, $i = 0, 1, \dots, n$, 当 $x_i = 1$ 表示基因是被表达, 当 $x_i = 0$ 表示基因是未被表达, F_i 表示在基因编码相互作用中影响像素点 i 的可能反映规则, 对于每一个 $f_j^{(i)}: \{0, 1\}^n \rightarrow \{0, 1\}$ 通过基因的影响决定像素点 i 的基因编码值以区分是否为图像轮廓特征, $l(i)$ 是影响像素点基因 i 可能的函数的个数^[9], 如果有基因 i 个, 则能够得到的基因结构个数为 $N = \prod_{i=1}^n l(i)$ 。

假如某个图像有25个元素, 则实现基因编码只需要2个基因, 第一个基因由 $\{0, 1, 2, 3, 4\}$ 组成, 第二个基因由 $\{A, B, C, D, E\}$ 组成, 通过数据组合实现^[10]。

则节点基因 i 被选中的概率为: $P_i = \prod_{j=1}^n c_{kij}^{(j)}$

其中, j 代表基因列; k 为选择矩阵。

个体的适应值函数为: $f_i = \begin{cases} n, & n > c_p \\ 0, & \text{else} \end{cases}$

式中, c_p 为具有较多样本的类的适应度样本的个数; n 为正确适应的个数。

个体绝对误差的计算为: $E_{ij} = |P_{ij} - T_j|$

式中, P_{ij} 为第 i 个个体基因对第 j 个特征信息的预测值; T_j 第 j 个特征信息的目标值; E_{ij} 为第 i 个

体基因对第 j 个适应度函数的绝对误差。 E_{ij} 越小,说明第 i 个个体基因越接近第 j 个特征信息的目标值,本文选择的 E_{ij} 为 0.00487。

2.3 图像轮廓特征信息提取研究过程

图像轮廓特征信息提取可规划为评价指标最优化的过程^[11],把图像信息分为 2 类:特征信息和非特征信息,这分别对应了分类中的 2 分类问题。图像像素视为每个基因,而每个基因所编码的属性各不相同,通过此对图像的相关属性进行筛选,即自动进行特征提取。假设图像轮廓特征信息模型的提取样本 s_i 的格式: $s_i(\alpha_1, \alpha_2, \dots, \alpha_n, c)$ 。

其中,下标 i 代表记录号, $\alpha_k (1 \leq k \leq n)$ 代表像素属性 k 经过预处理后得到的值, c 是类标号(即特征信息、非特征信息类标号)代表该记录素属性 k 所属的类别,分别为特征信息类(P 类)和非特征信息类(N 类)。

算法实现过程:

1. 输入图像,初始化基因;
2. 根据图像像素个数选择基因数量进行编码,计算基因个体的适应值函数和目标函数对像素轮廓特征信息进行提取;
3. 比较判别个体绝对误差,符合条件终止则进行步骤 4,否则转到步骤 2;
4. 输出结果。

3 实验仿真

图像轮廓特征信息提取所用程序为 Matlab 7.0, CPU 为 3.4 GHz, 内存 1024 MB, 操作系统 XP, 硬盘为 SATA2 接口, 用蒙特卡罗仿真 30 次对比实验效果。对图像轮廓特征信息进行提取, 其中在原始图像图 3.1 中加入高斯噪声(均值为 0, 方差为 0.01)如图 3.2 所示, 检测的视觉结果如图 3.2、图 3.3、图 3.4、图 3.5、图 3.6 所示。



图 3.1 原始图像



图 3.2 原始图像加入高斯噪声



图 3.2 Fourier 变换法提取结果



图 3.3 边界方向直方图法提取结果



图 3.4 粒子群算法提取结果



图 3.5 形态学理论提取结果



图 3.6 基因编码提取结果

从图 3.2 Fourier 变换法提取结果、图 3.3 边界方向直方图法提取结果、图 3.4 粒子群算法提取结果、图 3.5 形态学理论提取结果、图 3.6 基因编码提取结果对比中可以看出,基因编码提取结果对图像轮廓边缘很清晰,抗噪声能力强,同时没有伪边缘,这是因为基因编码算法通过反映规则对像素点的基因编码值以区分是否为图像轮廓特。其他算法受噪声影响较大,检测到的边缘宽度较粗,定位精度不高。

为了给出不同算法的时效性,表 1 给出了不同算法的处理时间。

表 1 不同算法处理时间

算法	Fourier 变换法	边界方向直方图法	粒子群算法	形态学理论	基因编码算法
处理时间	5.3493s	4.1825s	3.4234s	3.7980s	1.4837s

从表 1 的处理时间中可以看出,基因编码算法处理时间最少,这是因为基因编码算法在对图像像素进行编码的时候是根据像素总数选择不同的基因数量,不同基因的长度则数据组合起来相差比较大,因此在处理的时候,选择基因的长度相对较长,组合的数据多代表了图像全部像素。

为了给出不同算法的定位性能,表 2 给出了不同算法在检测出的边缘点保持指数(EPI),计算为:

$$EPI = \frac{\sum |P_s - P_{sn}|}{\sum |P_o - P_{on}|}$$

其中, P_s 为去噪后的像素值; P_{sn} 为 P_s 的相邻像素; P_o 为原始图像的像素值; P_{on} 为 P_o 的相邻像素。EPI 其值越接近 1 的边缘保持能力越强。

表 2 不同算法检测出的 EPI 结果

算法	Fourier 变换法	边界方向直方图法	粒子群算法	形态学理论	基因编码算法
EPI	0.4387	0.63289	1.2991	1.4387	0.9355

从表2的不同算法检测出的EPI结果中得出图像轮廓信息检测定位精度较高,基因编码算法对像素进行被表达和未被表达分类,只要是未被表达分类的像素,基因编码算法就认为是非轮廓曲线。

4 总结

用CEP算法对像素映射基因个体表达式,当基因是被表达时则像素点的基因编码值区分为图像轮廓特征,根据每个基因所编码的属性各不相同,通过此对图像的相关属性进行筛选进行特征提取,得出基因编码提取的图像轮廓边缘很清晰,抗噪声能力强,时效性快,边缘点保持指数接近1。未来基因编码若能够实现对像素进行二次编码,那么图像轮廓信息提取的精度将更加提高。

参考文献:

- [1] Chang Danhua, Zhao Dan, Yang Fengming, et al. Improvement of fisheye image contour extraction and image correction algorithm[J]. Laser & Infrared, 2012, 42(4): 448-451. (in Chinese)
常丹华,赵丹,杨峰明,等.提取鱼眼图像轮廓的算法改进及图像校正[J].激光与红外,2012,42(4):448-451.
- [2] Yang Dan, You Lei, Zhang Xiaohong, et al. Fisheye image contour extraction algorithm based on region growing[J]. Computer Engineering, 2010, 36(8): 217-218, 221. (in Chinese)
杨丹,游磊,张小洪,等.基于区域生长的鱼眼图像轮廓提取算法[J].计算机工程,2010,36(8):217-218,221.
- [3] Yuan Da, Zhang Caiming, Li Jinjiang, et al. A high-accuracy contour extraction algorithm for MR images based on Mumford-Shah model[J]. Chinese journal of computers, 2009, 39(2): 268-274. (in Chinese)
原达,张彩明,李晋江,等.基于Mumford-Shah模型的高精度MR图像轮廓提取算法[J].计算机学报,2009,39(2):268-274.
- [4] Shao Mingsheng, Wang Qihua. Blurred image restoration based on frog leaping algorithm[J]. Laser & Optoelectronics Progress, 2012, 49(2): 0210031-0210036. (in Chinese)
邵明省,王其华.基于蛙跳算法的模糊图像复原[J].激光与光电子学进展,2012,49(2):0210031-0210036.
- [5] Zhu Mingfang, Tang Changjie, Chen Anlong, et al. Mining compact function based on naive gene expression programming[J]. Journal of University of Electronic Science and Technology of China, 2010, 39(2): 284-288, 310. (in Chinese)
朱明放,唐常杰,陈安龙,等.基于朴素基因表达式编程挖掘紧致函数[J].电子科技大学学报,2010,39(2):284-288,310.
- [6] Zhang Jianbin, Zhu Mingfang, Ren Yanling. Business total prediction of posts and telecommunication based on gene expression programming[J]. Computer Engineering and Design, 2010, 31(18): 4124-4127. (in Chinese)
张建斌,朱明放,任艳玲.基于基因表达式编程的邮电业务量预测[J].计算机工程与设计,2010,31(18):4124-4127.
- [7] Peng Yuzhong, Yuan Changan, Mai Xiongfa, et al. Survey on theoretical research of gene expression programming[J]. Application Research of Computers, 2011, 28(2): 413-419, 438. (in Chinese)
彭昱忠,元昌安,麦雄发,等.基因表达式编程的理论研究综述[J].计算机应用研究,2011,28(2):413-419,438.
- [8] Xiang Yong, Tang Changjie, Zhu Mingfang, et al. Embedded gene expression programming and its application in function mining[J]. Journal of University of Electronic Science and Technology of China, 2011, 40(1): 116-121. (in Chinese)
向勇,唐常杰,朱明放,等.内嵌基因表达式编程及其在函数发现中的应用[J].电子科技大学学报,2011,40(1):116-121.
- [9] Xu Honglin, Wang Shitong. Structure analysis of probabilistic boolean networks for gene regulatory networks modeling[J]. Journal of Jiangnan University: Natural Science Edition, 2004, 3(5): 465-470. (in Chinese)
徐红林,王士同.用于基因调节网络建模的概率布尔型基因网络结构分析[J].江南大学学报:自然科学版,2004,3(5):465-470.
- [10] Zhang Guowei, Shao Shihuang, Zhang Ying, et al. Boolean genetic networks and its recent research[J]. Journal of Donghua University, 2006, 32(2): 127-129, 134. (in Chinese)
张国伟,邵世煌,张颖,等.布尔基因网络及其研究[J].东华大学学报:自然科学版,2006,32(2):127-129,134.
- [11] Liu Kaiying, Wu Jiang, Li Taiyong. Credit evaluation based on bagging GEP classifier integration[J]. Computer Engineering, 2011, 37(8): 213-215. (in Chinese)
刘凯英,吴江,李太勇.基于装袋GEP分类器集成的信用评估[J].计算机工程,2011,37(8):213-215.