

⑦

115-118

曲线轮廓汉字的两个关键技术的研究

刘东林 杨功能[✓] 穆玉杰

TP391

(西北大学数学系, 710069, 西安, 第一作者 25 岁, 女, 硕士生)

摘要 主要获得了以下结果: ①在型值点的提取技术中, 提出了新的型值点的整理排序方法, 此算法在计算机上得以实现; ②解决了复杂汉字的填充方向问题, 并给出算法, 在计算机上实验有效。

关键词 型值点; 曲线轮廓汉字; 填充

分类号 TP399; H122

汉字处理

字形技术

汉字是以形成字, 每个汉字都可看作一个图形, 与西文相比较, 汉字本身具有字形复杂、数量大、字体多等特点。因此为了使汉字描述美观流畅且真实地再现原字的风格, 又能大量压缩所需存储的信息量, 从而采用曲线轮廓来描述汉字的方法成了 90 年代的必然趋势。

曲线轮廓汉字是当前较新的一种方法, 它不仅能准确地把字稿信息描述下来, 保证了还原字形的质量, 而且对数据进行了大量的压缩, 使用起来可以任意放大、缩小及花样变化, 基本上满足了电子印刷中字形质量的要求。

字形技术是汉字处理的主要部分, 主要包括型值点的提取技术、还原技术和填充技术。

下面将对曲线轮廓汉字的型值点的提取技术和复杂汉字的填充方向问题做一讨论和研究。

1 型值点的提取技术

要得到一个曲线轮廓汉字, 首先必须具有这个汉字轮廓上点的信息。这些提取出来的点称为此汉字的型值点。型值点的提取应达到少而精, 即型值点的数目尽量少, 因为这直接影响着汉字的存储量, 但这些型值点必须确保还原后的字形质量, 美观且不失真。

型值点的提取方法通常采用向量法, 所谓向量法, 即沿原轮廓线上依次搜索判断提取, 主要分为二步进行。

第一步, 对同一直线上的点筛选提取。

第二步, 采用曲率值的方法对上步提取出来的型值点做第二次精选, 得到最后所需的型值点的信息。

值得注意的一点是, 为了保证型值点的提取沿同一方向顺序下来, 在此不妨设为顺时针方向, 则对第一个型值点, 要得到它的下一个型值点, 对它的上、下、左、右、右上、右下、左上、左下 8 个方向的搜索方向定为顺时针方向, 如图 1 所示。

由于原字模库的原因造成汉字轮廓在形成过程中的不连贯, 即一条封闭的轮廓曲线分成几条开曲线来描述。如人字, 就分为图 2 所示的两段来提取。这样提取出来的型值点不仅大大影响了曲线轮廓汉字的还原速度, 而且使填充工作难以实现。

为了弥补这一不足, 就需对所提取出来的型值点进行重新整理排序连接。首先将型值点按断点来分

组存放。所谓断点,即若一个型值点的 8 个搜索方向上不再有原轮廓线上的点,则称这个型值点为断点。在每个断点后面插入坐标值(0,0),从而型值点被坐标值(0,0)分为一组,而坐标值(0,0)称为它所在这一组型值点的标志值。

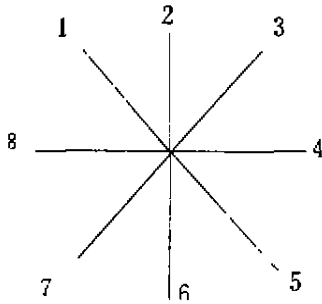


图 1 8 个方向的搜索顺序

Fig. 1 Searching Sequence of Eight Direction



图 2 分为两段开曲线提取的人字

Fig. 2 Two Not-closed Curves of Ren Character

然后可对这些型值点组整理排序连接。设每一断点的 X, Y 坐标值存放在 AX, AY 数组中,而每组型值点的第一个型值点的 X, Y 坐标值存放在 BX, BY 数组中。

在此之前,先介绍两个定义:正常断点及非正常断点。

正常断点:这个断点所在的型值点组能生成一条闭曲线。

非正常断点:这个断点所在的型值点组不能生成一条闭曲线。

用数学公式,正常断点与非正常断点的定义又可描述为:

对第 i 组型值点的断点,给出下式:

$$\text{abs}(AX[i]-BX[i]) < d \&\& \text{abs}(AY[i]-BY[i]) < d \quad (1)$$

若这个断点满足公式(1),则称它为正常断点。反之,为非正常断点。

设有 n 组型值点,对它们进行排序连接的具体步骤如下:

- 1) 利用公式(1)筛选出具有非正常断点的型值点组。

2) 对这些有非正常断点的型值点组进行判断,修改它们的标志值:

若第 i 组型值点与第 j 组型值点尾首相连, ($i < j$),即满足公式(2),则将第 i 组型值点的标志值修改为(0, j)。

$$\text{abs}(AX[i]-BX[j]) < d \&\& \text{abs}(AY[i]-BY[j]) < d \quad (2)$$

若第 i 组型值点与第 j 组型值点尾尾相连, ($i < j$),即满足公式(3),则将第 i 组型值点的标志值修改为(1, j)。

$$\text{abs}(AX[i]-AX[j]) < d \&\& \text{abs}(AY[i]-AY[j]) < d \quad (3)$$

d 为计算机上反复实验得一较小的值。在此 d 值取为 2。

3) 根据每一组型值点的标志值将它们排序连接,这样保证了每一组型值点均生成一条闭曲线,且每条曲线的生成方向均为顺时针方向。

经过重新整理排序,减少了型值点组的数目,不仅提高了曲线还原速度,而且十分有利于曲线轮廓汉字的填充。

此算法已在计算机上予以实现,实践证明是有效且成功的。程序框图如图 3:

设对已有的 n 组型值点进行重新整理排序,修改后的型值点的信息存放在 rrx, rry 数组中。

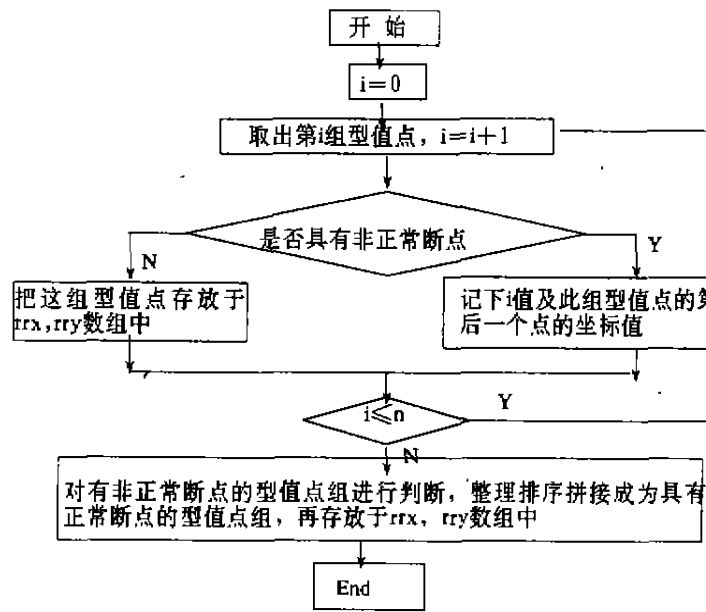


图 3 型值点的整理排序程序框图

Fig. 3 Flowchart About Chosing Technology of Data Points

2 复杂汉字的填充方向问题

这里所说的复杂汉字不是指字形结构复杂的汉字,而是指那些填充复杂的汉字,也就是有非独立轮廓笔画的那些汉字。在这里将每一组型值点生成的一条闭曲线,即一个封闭轮廓称为一个轮廓笔画。

非独立轮廓笔画是指这一轮廓笔画包含在其他轮廓笔画的内部,或者它的内部包含有其他的轮廓笔画。

若一个汉字的每一轮廓笔画都是相互独立的,即互不包含。而经过重新整理排序后的型值点保证了生成一条闭曲线,且生成方向为顺时针方向,那么每一轮廓笔画的填充方向很容易确定为内法线方向。

而对于有非独立轮廓笔画的汉字,如口字,有内外两个轮廓笔画,所需填充的部分为两个轮廓笔画的中间部分,显然内轮廓笔画的填充方向不再为内法线方向。若简单地将内、外轮廓笔画的填充方向都定为内法线方向,则会造成如图 4 所示的不正确的填充结果。

因此对曲线轮廓汉字填充前,必须确定它的每一轮廓笔画的填充方向。为了解决这一问题,首先给出曲线轮廓汉字的如下两个性质。

性质 1 任意两个轮廓笔画都不相交,它们或者相互包含,或者相互独立。

性质 2 若轮廓笔画 i 包含在轮廓笔画 L_1, \dots, L_n 中,令 $L_k = \max\{L_1, \dots, L_n\}$, ($i > L_1, \dots, i > L_n$) 则轮廓笔画 i 的填充方向与轮廓笔画 L_k 的填充方向相反。

由以上两个性质,即可以判断确定每一轮廓笔画的填充方向,具体步骤如下:

设有 n 组型值点生成 n 条闭曲线,也即 n 个轮廓笔画 L_1, \dots, L_n 。

1) 求出包含每个轮廓笔画的最小 BOX,即求出每组型值点的最大 X, Y 坐标值,记为 X_{\max}, Y_{\max} ,及最小 X, Y 坐标值,记为 X_{\min}, Y_{\min} 。以点 (X_{\max}, Y_{\max}) 及点 (X_{\min}, Y_{\min}) 为对角点生成的矩形为含轮廓笔画 i 的最小 BOX,记为 RL_i 。

2) 对轮廓笔画 i 的最小 BOX RL_i ,在 $\{RL_1, \dots, RL_{i-1}\}$ 中寻找包含它的最小 BOX,若存在,记为 RL_j ,再采用累积交点的方法判断是否包含在轮廓笔画 L_j 中,若包含,则轮廓笔画 L_i 的填充方向与轮廓笔画 L_j 的填充方向相反。若不包含,则轮廓笔画 L_i 的填充方向为内法线方向。若不存在,则轮廓笔画 L_i 的填

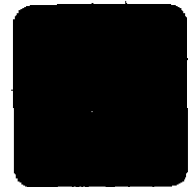


图 4 填充方向均取为内法线方向的填充结果
Fig. 4 The Wrong Resultes Caused by Wrong Filling-direction

充方向为内法线方向。

3) 对曲线轮廓汉字的每一轮廓笔画得到了它的填充方向, 则可对这个汉字进行正确的填充。

以下程序框图如在上述实践证明有效, 可成功地得到每一轮廓笔画的填充方向。该程序框图解释如下: 设有 n 组型值点, 生成 n 条闭曲线 L_1, L_2, \dots, L_n , S 数组存放 $L_i, i = 1, \dots, n$ 的填充方向的标志值。

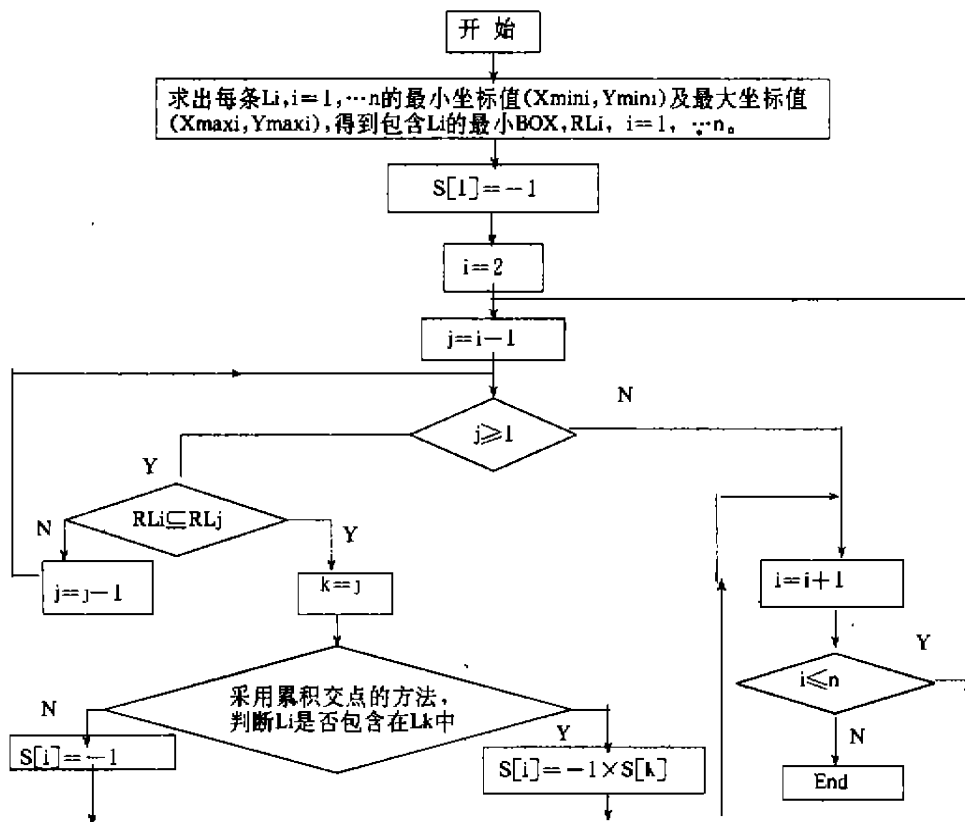


图 5 确定“口字型”填充方向的程序框图

Fig. 5 Flowchart About Filling-direction of "Kou Character form"

参 考 文 献

- 1 唐荣锡, 汪嘉业, 彭群生等. 计算机图形学教程. 北京: 科学出版社, 1990
- 2 彭寿全, 黄可. 汉字信息处理. 成都: 电子科技大学出版社, 1994
- 3 施法中. 计算机辅助几何设计与非均匀有理 B 样条. 北京: 北京航空航天大学出版社, 1994
- 4 Gerald Farin. Curves and Surface for Computer Aided Geometric Design. Academic Press, INC 1990
- 5 孙家广, 杨长贵. 计算机图形学. 北京: 清华大学出版社, 1995

责任编辑 张素敏 曹大刚

A Study on Two Key Technology of Curves Outline Character

Liu Donglin Yang Gongneng Mu Yujie

(Department of Mathematics, Northwest University, 710069, Xi'an)

Abstract Mainly the two following results are provided: ① A new algorithm about choosing technology of data points is presented, and turn it true on computer; ② The filling direction of difficult Chinese Character is solved, and again fine results on computer.

Key words data points; cover outline chinese character; fill