

点阵汉字向量化算法

毛自强 朱耀庭

(南开大学)

摘要

本文提出了点阵汉字向量化的方法,解决了在笔绘图仪上输出汉字的问题。此方法采取了点分类和有选择的搜索象点邻点,根据笔划书写特征建立了选择规则。本文给出了通道搜索算法和通道回溯算法,开发了两个实用的软件模块。最后举出了在 SR-6602 六笔绘图仪上画出的字例。

一、向量化方法

计算机自动绘图中,经常需要用笔绘图仪标绘出笔划汉字。本文提出一种软件实现方法,解决了笔绘图仪写汉字问题。这个方法就是通过点阵图形向量化^[1],把点阵汉字转化为笔划汉字。

点阵汉字图形是由 $N \times M$ 的象点矩阵构成的。为搜索方便,扩充此阵为 $(N + 2) \times (M + 2)$, 记为 A , 仍称之为象点矩阵。 I 行 J 列的象点用 (I, J) 表示, 该点值记为 $A(I, J)$ 。它可取 1 或 0, 并设定: $A(K, J) = 0, (K = 0, N + 1)$; $A(I, L) = 0, (L = 0, M + 1)$ 。其中 $I = 0, 1, \dots, N + 1, J = 0, 1, \dots, M + 1$ 。

本文仅讨论单线体汉字, 以下用 P 记扩充前象点矩阵中的一个任意的象点, 值记为 $A(P)$ 。

P 的邻点是指与 P 相邻的八个点 $P_i, (i = 1, \dots, 8)$, 见图 1。 P_i 点相对于 P 所在的方向称为 P 的第 i 个方向。

通道是指一串象点 $Q_1, \dots, Q_s (s \geq 2)$ 连成的折线, 其中 $A(Q_i) = 1, (i = 1, \dots, s)$, 且对所有 $i < s$, Q_{i+1} 是 Q_i 的邻点。

若象点 P 的值为 1, 且 P 的第 1、3、5、7 方向上非零值象点个数大于 2, 或者 P 的第 2、4、6、8 方向上的象点值均非零, 则称该点为交叉点。

汉字点阵图形向量化是把点阵转化为通道, 并适当地选取一部分通道, 使之构成笔划汉字。

现在对象点矩阵从上到下、从左至右进行搜索。根据汉字笔划书写特征, 当前点 P 相

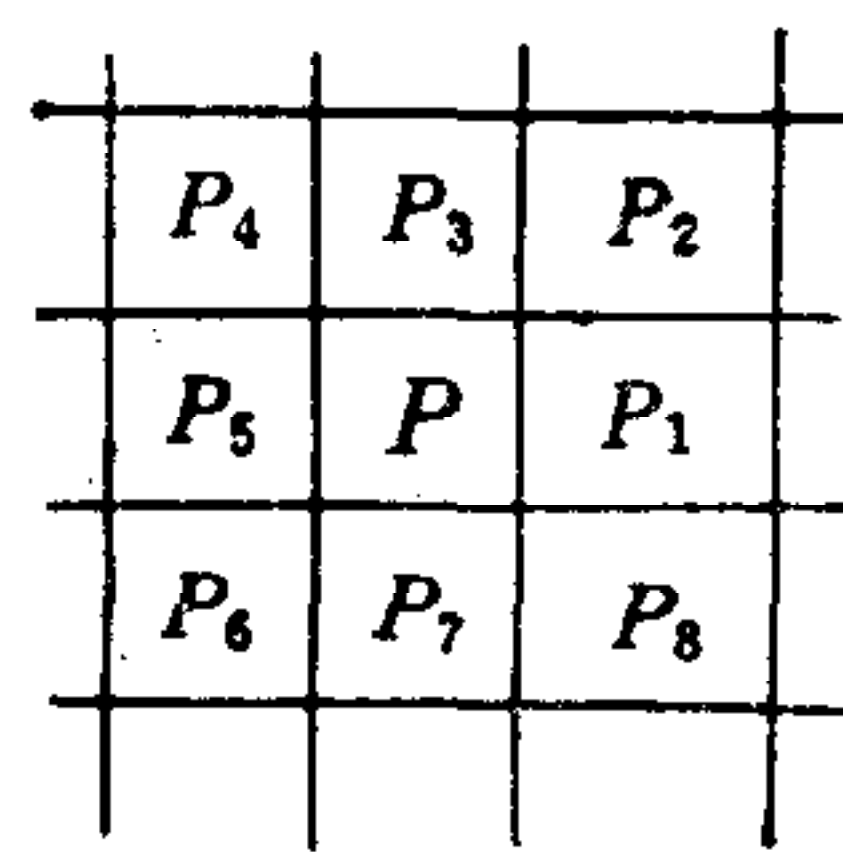


图 1. P 的邻点

对于前一个点的方向决定了要被搜索的邻点和这些邻点的顺序,表 1 给出了一种选择规则。其中 $order(0)$ 为搜索时当前点相对于前一个点的方向, $order(I)$ 为第 I 个要被搜索的邻点相对于当前点的方向。当 $order(I)$ 为 0 值时,即没有第 I 个要被搜索的邻点,故对当前点邻点的搜索结束。这里 $I = 1, \dots, 5$ 。在搜索一通道的初始点时,规定初始点的前一个点在初始点的第 5 个方向上。

搜索过程中,为保证汉字字形的完整,对搜索中的初始点和终止点以及交叉点 P ,置

表 1 选择搜索规则

$order(0)$	1	2	3	4	5	6	7	8
$order(1)$	6	6	7	8	1	2	0	4
$order(2)$	7	5	1	7	7	1	0	0
$order(3)$	0	7	6	1	8	4	0	0
$order(4)$	0	8	8	6	6	0	0	0
$order(5)$	0	0	2	0	2	0	0	0

$A(P) = 2$, 以供再搜索;其它值为 1 的象点 P , 置 $A(P) = 3$, 以保证不再被搜索到。

此外,采用了两个方法来提高字形的质量,一个方法是对 A 的每行进行两次搜索,另一个是用通道回溯搜索^[2]。在搜索到一段通道的终止点后,逐步回溯到最近的交叉点,向另外的通道分枝搜索,最后回溯到初始点。

二、通道搜索算法

设 (PI, PJ) 为当前点 P , (I', J') 为 P 的邻点, A 为扩充后的象点矩阵。通道搜索算法如下:

1 $I = 0, JJ = 0$;

2 $I = I + 1, JJ = 0$; 若 $I = N + 1$, 则结束;

3 $JJ = JJ + 1, J = [(JJ - 1)(\text{mod}M)] + 1$; 若 $JJ = 2M + 1$, 则 GOTO 2;

4 若 $A(I, J) = 1$ 或 2, 则 GOTO 5; 否则, GOTO 3;

5 置 $order(0) = 5, PI = I, PJ = J$, 抬笔移动到 (PI, PJ) 点, 落笔, 记 $A(PI, PJ) = 2$;

6 根据 $order(0)$ 的值确定搜索的邻点及其顺序; 若 (PI, PJ) 为交叉点, 则记 $A(PI, PJ) = 2$;

7 按已确定的邻点及其顺序搜索, 若存在邻点 (I', J') , 使 $A(I', J') = 2$, 则落笔移动到 (I', J') , 并根据 (I', J') 相对于 (PI, PJ) 的方向对 $order(0)$ 重新赋值, 记 $PI = I', PJ = J', A(I', J') = 3$, GOTO 6; 否则, GOTO 8;

8 按已确定的邻点及其顺序搜索, 若存在邻点 (I', J') , 使 $A(I', J') = 1$, 则落笔移动到点 (I', J') , 且根据 (I', J') 相对于 (PI, PJ) 的方向, 对 $order(0)$ 重新赋值, 记 $PI = I', PJ = J', A(I', J') = 3$, GOTO 6; 否则, 记 $A(PI, PJ) = 2$, GOTO 3。

三、通道回溯算法

$(PI(K), PJ(K))$ 和 $order(L, K)$ 分别表示搜索某通道第 K 步时象点和它周围要被搜索的邻点序号, $order(0, K)$ 为 $(PI(K), PJ(K))$ 相对于其前一个点的方向号, (PPI, PPJ) 为当前点. 通道回溯算法如下:

- 1 $I = 0, J = 0$;
- 2 $I = I + 1, J = 0$; 若 $I = N + 1$, 则结束;
- 3 $J = J + 1$, 若 $J = M + 1$, 则 GOTO 2;
- 4 若 $A(I, J) = 1$, 则 GOTO 5, 否则 GOTO 3;
- 5 $order(0, 0) = 5, PI(0) = I, PJ(0) = J, PPI = I, PPJ = J, K = 0$, 抬笔移动到 (I, J) , 落笔, 记 $A(I, J) = 2$;
- 6 根据 $order(0, K)$ 的值确定搜索的邻点及其顺序, 若 (PPI, PPJ) 是交叉点, 记 $A(PPI, PPJ) = 2$;
- 7 按已确定的邻点和顺序搜索, 若存在邻点 (I', J') , 使 $A(I', J') = 1$, 则落笔移动到 (I', J') , 置 $K = K + 1$, 根据 (I', J') 相对于 (PPI, PPJ) 的方向决定新的 $order(0, K)$ 值, 记 $PPI = I', PPJ = J', A(PPI, PPJ) = 3, PI(K) = PPI, PJ(K) = PPJ$, GOTO 6; 否则, GOTO 8;
- 8 $K = K - 1$, 若 $K \leq 0$, 则 GOTO 3; 否则, 若 $A(PI(K), PJ(K)) = 2$, 则抬笔移动到 $(PI(K), PJ(K))$, 且 GOTO 7; 否则, GOTO 8;

四、向量化模块

将上述算法编成程序模块, 在 IBM PC/XT 上对 CCBIOS 实现了向量化. 只要指

插 叉 荏 茶 查 碓 捺 岔 差 詫 拆 柴 豺 撵 掺
 铲 产 阐 颤 昌 猖 场 尝 常 长 安 俺 按 暗 岸
 凹 敖 熬 翱 袄 傲 奥 懊 澳 喀 咖 卡 咯 开 揩
 楷 凯 慨 刊 堪 勘 坎 砍 看 康 糠 扛 抗 亢 炕
 买 麦 卖 迈 脉 瞞 蛮 满 芒 茫
 哪 呐 钠 那 娜 纳 乃 奶 耐 奈

图 2 通道搜索算法模块的字例

啊阿唉扎渣札轧铡闸眨栅榨咋乍
 炸诈摘斋宅窄债寨瞻毡詹粘沾盍斩
 芭捌扒叭吧芭八疤巴拔跋靶把耙坝
 霸罢白柏爸百摆佰败拜稗斑班搬扳
 帆番翻帆繁凡烦反该改
 航昂盎哈骸孩海亥害咳
 酣憨邯邯单单含涵函喊憾憾

图3 通道回溯算法模块的字例

明汉字及其大小,位置和方向等参数,就可以在 SR-6602 笔绘图仪上绘出汉字(见图 2、图3)。

向量化模块结构如下:

- (1) 在点阵汉字库中寻找汉字的内码;
- (2) 由内码计算出字模信息所在的地址;
- (3) 取出字模信息,实行向量化转换。

提出的算法是一个在线实施过程,占用的存储空间小,实现速度快,能满足实时绘图要求。

感谢张朝池副教授的热情支持和耐心指导。

参 考 文 献

- [1] Gibson, L., Lucas, D., Vectorization of Raster Images Using Hierarchical Methods, CGIP 20 (1), Sep. 1982, 82—89.
- [2] 何华灿,人工智能导论,航空专业教育编审组,1983.

VECTORIZATION ALGORITHMS FOR THE DOT MATRIX CHINESE CHARACTER

MAO ZIQIANG ZHU YIAOTING

(Nankai University)

ABSTRACT

The vectorization method for the dot matrix Chinese character is proposed in this paper. The problem of Chinese character output by pen-plotters has been solved. Dot sorting and selective search of the adjacent dots are adopted. The selective search rule has been established according to the features of the stroke writing. Two search algorithms of vectorization, the Path Search Algorithm and the Path Backtracking Algorithm, are given, and two useful modules are developed. Finally, the examples plotted by the SR-6602 Plotter are given.