No. 2 Apr. ,2004

## 汉字笔顺识别的算法与实现

## 李正华

(湖南城市学院信息与计算机科学系,湖南 益阳 413000)

摘 要:提出了一种点阵汉字的笔顺字模的描述方法以及基于这种描述方法的汉字笔顺识别的算法和实现。

关键词:汉字笔顺识别;识别算法;点阵汉字

中图分类号:TP317.2

文献标识码:A

文章编号:1008-1666(2004)02-0039-02

# Algorithm and implementation in Chinese Character's Order of Strokes Recognition

## LI Zheng-hua

(Department of Information & Computer Science, Hunan City University, Yiyang, 413000, China)

Abstract: On the basis of the coding principle about Chinese characters' order of strokes, the algorithm and implementation in Chinese characters' order of strokes recognition suggested in the paper.

Key words: Chinese characters' order of strokes recognition; algorithm of recognition; coding principle

## 1 引言

汉字笔顺是汉字书写规范化的重要环节。以显示汉字笔顺为主要功能的汉字笔顺字典软件的关键技术在于生成汉字笔顺字库。所谓汉字笔顺字典就是以动画形式模拟汉字书写,同时还显示每一笔的状态和该汉字的笔画数,这种字典无论用于数学还是作个人查阅都非常实用。通过对汉字结构的分析和笔顺显示的要求,提出一种汉字笔顺的识别算法。该算法只需少许人工干预即可提取汉字的笔顺信息,且适用于任何大小任何字体的点阵汉字。

## 2 问题描述

国标点阵字模用一个二进制位表示一个点的显示与否。一个 24 \* 24 点阵的汉字字模需要 3 \* 24 = 72 个字节来表示,因为每行 24/8 = 3 个字节,共 24 行,其特点是比特位所在位置即为点的位置。显示扫描时只要将对应字模按从左到右、从上到下顺序进行即可,也就是说字模的存储位置包含了点的显示位置信息。这种字模信息的描述方式,对点阵汉字的显示而言是精简而完善的,但是它没有也不可能包含汉字的笔顺信息。

要利用点阵字模模拟笔顺写字,只要将相应笔画的显示点按笔顺顺序显示即可。因此,笔顺点阵字模必须包含两种信息,即显示点的位置信息和它的笔画顺序信息。

### 3 算法描述

显然,根据上面叙述的笔顺点阵字模要求,完全可以按照已有的点阵字模用人工编码实现笔顺点阵字模,但工作量大、容易出错。下面先说明笔顺字模的描述规则,然后给出从点阵字模生成笔顺字模的汉字笔顺识别算法。

#### 3.1 笔顺字模的描述

总的原则是将点阵字模一个比特位表示一个点扩充为一个字节表示一个点。原因有二:其一,必须用一个字节表示一个点才能包含该点的笔顺信息;其二,笔顺字模在字库中的索引方法上与点阵字模的基本一致,只是偏移量扩大8倍。类似于点阵字模,该笔顺字模中点的存储位置决定了点的显示位置,点的字节值(ASCII 字符的值)则包含点的笔顺信息。点的字节值的确定规则:

- (1)不显示点即空白点用 ASCII 字符的空格表示。
- (2)笔画点用从 ASCII 字符的'A'开始依次表示,而每一笔的起点用从 ASCII 字符的'S'开始依次表示。例如,

任何一个字第一笔的起始点用'S'表示,该笔的其它点均用'A'表示。

(3)两笔交叉的点即公共点用 ASCII 字符的'8'表示。 注:汉字中不存在三笔交叉。

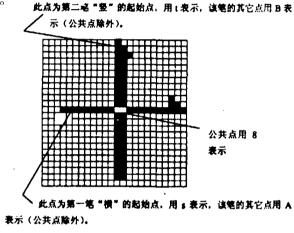


图 1 "十"字笔顺字模的描述

#### 3.2 算法描述

本识别算法分两个阶段实现。

第一阶段是准备阶段:

- (1) 把要识别的汉字如图 1 显示以备操作。
- (2) 将该点阵字模的数据改造后保存在一个一维数组中。数组元素的顺序与点阵字模的比特位顺序——对应即点阵字模的第1至第3个字节对应数组的第1至第24个元素,依此类推;比特0对应的数组元素值是0x20(表示该点不显示,称为空白点),比特1对应的数组元素值是0x40(表示该点显示,称为原始点)。这样可以利用数组元素的下标标记点的位置。

#### 第二阶段是笔顺识别阶段:

- (1)本笔画的公共点和边界点。所谓边界点是指从属本笔但和其它笔画相邻接的点。
  - (2)标识本笔画的起始点。
  - (3)开始笔画识别。步骤如下:

Step1 初始化种子点队列 Q1 和公共点队列 Q2。种子点是指将要以之为基准进行相邻搜索的笔画点。队列 Q1 和 Q2 的元素值均为该点对应的数组元素的下标即该点的位置。

Step2 起始点人队列 Q1 并将该原始点标记为本笔点。根据 3.1 可知,本笔点的值  $\approx$  A 的 ASCII 码值 + k - 1,

k表示本笔是第 k 笔。

Step3 种子点出队列。如果队列为空,则本笔画识别 完毕,转至 Step6。

Step4 以种子点为基准搜索相邻的6个点(如果种子点处在点阵边缘则相邻点少于6个,因此要进行判断)。若相邻点是原始点则将其标识为本笔点并入队列Q1;若相邻点是边界点则只将其标识为本笔点,边界点不会成为种子点,因此不将其人队列Q1;若相邻点是公共点则将其标识为本笔点并入队列Q1和Q2。

Step5 重复 Step3 和 Step4。

Step6 将公共点队列 Q2 中的所有公共点恢复为原始点以备另一笔使用。

重复 Step1 至 Step6,直至所有笔画识别完毕。

## 4 算法实现

用 Visual C++ 6.0 实现了汉字识别算法。识别过程中的人工干预用鼠标能轻松完成。为使识别效果简单明了,识别出的每一笔均采用不同颜色显示。识别完成后可将生成的笔顺字模保存在笔顺字库中备用。图 2 为汉字笔顺识别程序对"匹"字的识别。

调用笔顺字模,采用类似算法即可逐点显示模拟汉字 规范的笔顺书写过程。

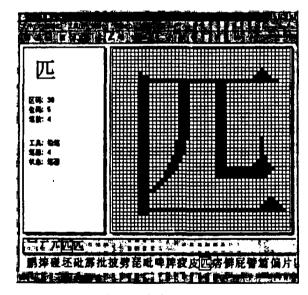


图 2 汉字笔顺识别程序

## 5 结 论

采用本算法生成的汉字笔顺识别程序,适用于任何大小任何字体点阵汉字的笔顺识别,平均每分钟识别3个汉字的笔顺。

## 参考文献:

- [1]李正华. 汉字笔顺字典软件[J]. 计算机应用与软件,2003,(5): 08-123.
- [2]朱巧明、汉字信息处理基础[M]. 北京:清华大学出版社,1997.
- [3]刘金龙、Visual C++ 6.0 类参考详解[M]. 北京:清华大学出版社, 1999、