

汉字矢量字形输出的新途径

孙亚梅 孙以义

(南京大学大地海洋科学系)

提要 汉字矢量字形的输出是实现地图图面注记自动化的关键。汉字信息处理可采用用户程序级、操作系统级和硬件级三种方式。一、三两种方式已经过多年的实践,表明都有不尽人意之处。本文从操作系统级出发,将供显示和打印用的点阵式字模转化为适合绘图仪输出的矢量数据。这一方法可利用国标一、二级字库的汉字信息资源,并可以获得地图图面注记所需要的各种字体和字形的汉字。

关键词 汉字矢量字形 图面自动注记 中文磁盘操作系统 点阵式字模 字形栅格数据 字形矢量数据

机助地图制图当前已发展到了实用阶段,图面注记能否自动化便成了突出的问题。本文探求了一种图面注记自动化的新方法,本方法是在 IBM-PC 系列机上实现的,它基于现成的供打印和显示用的国标 GB2312-80 一、二级字库(含有 6763 个汉字,682 个图形符号和文字等),将库中的字形点阵信息转换成矢量笔划信息,驱动绘图仪实现自动定位和注记。

一、技术路径

近几年来,不少单位研制并建立了中文磁盘操作系统及其字库,为汉字矢量字形的输出提供了物质保证。

形成汉字矢量字形的主要技术路径是读取中文磁盘操作系统字库中的字形点阵信息,并将其转换成字形栅格数据,然后进行矢量化处理,得到字形矢量数据,再经字形变换,便形成所需要的汉字矢量字形(图1)。

(一) 字形栅格数据的获取

供打印输出的字库有三种类型: 简易型字库,一般为 16×16 以下点阵;普通型字库,通常是 24×24 或 32×32 的点阵;精密型字库,为 64×64 或 128×128 的点阵。点阵越大,字形越精美,但数据量就越大。对于 24×24 点阵的字形数据,每个汉字以 576 个黑、白格表示,计算机一个字节为 8 位,可以表示 8 个黑、白格。因此,每个汉字占 72 个字节,并以文件形式按区位码表排放,从字库中提取字形信息时,按 HZ\$ 计算:

$$\text{区码} = \text{ASC}(\text{HZ\$}) - 160$$

来稿日期: 1989 年 3 月。

$$\text{位码} = \text{ASC}(\text{RIGHT}(\text{HZ}\$1)) - 160$$

然后计算汉字在字库中的地址,即

$$\text{地址} = (\text{区码} - 16) \times 94 + \text{位码}$$

从此地址起,在字库中读取 72 个字节,即为某个汉字的字形信息。以二进制数“1”、“0”取代“黑”、“白”字形信息,便是所需要的字形栅格数据。

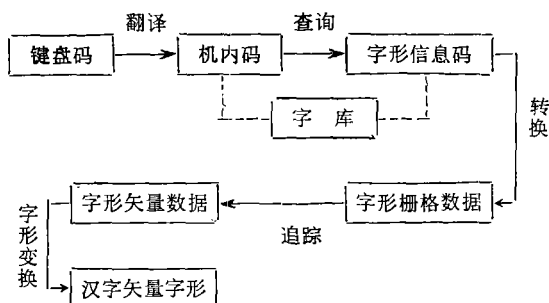
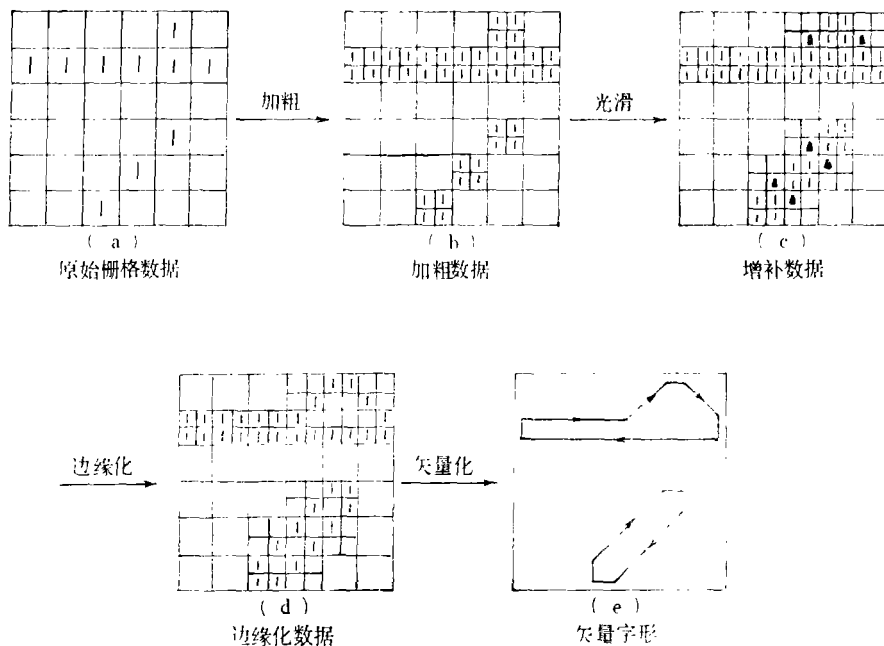


图 1 汉字矢量字形形成过程

Fig. 1 The flow chart of forming Chinese character in vector forms

(二) 矢量字形的输出

将所获得的字形栅格数据进行加粗与光滑、边缘化、矢量化和字形变换等处理,就可得到体态多样的字形矢量数据,驱动绘图仪便可输出矢量字形。



竖道: 点阵中的有值点 黑三角: 光滑增补点

图 2 字形处理

Fig.2 Processing of form of a word

1. 加粗与光滑 将原始栅格数据(图 2(a))的行和列各扩大 1 倍,使每个有值点变成 4 个有值点(图 2(b)),为之加粗。这样就可使笔划连续,且以双线封闭。加粗公式:

$$I_1 = 2I \quad I_2 = 2I - 1$$

$$J_1 = 2J \quad J_2 = 2J - 1$$

式中: I, J 为原始栅格数据中的有值点

$I_i, J_i (i = 1, 2)$ 为扩充栅格数据中的有值点。

光滑是在加粗笔划的转折处增补些点,将 4 联通的笔划变成 8 联通笔划,而使笔划的边缘柔和光顺,见图 2(c)。

2. 边缘化 这过程的实质是保留汉字笔划的轮廓,滤去笔划的内部点(内部点的特征是其上、下、左、右四方均为有值点,目的是为了获得空心字形,见图 2(d)。

PROCEDURE EDGE-DETECT

//边缘化算法//

IA: = 0

FOR I: = 1 TO 48 DO

FORJ: = 1 TO 48 DO

IZ: = BBT(I, J - 1) + BBT(I + 1, J) + BBT(I, J + 1) + BBT(I - 1, J)

IF IZ = 4 THEN IA: = IA + 1

X(IA): = I

Y(IA): = J

END IF

REPEAT

REPEAT

FOR I: = 1 TO IA DO

BBT (X(I), Y(I)): = 0

REPEAT

END EDGE - DETECT

3. 矢量化 这是汉字矢量字形输出的关键性过程。它包括将字形栅格数据转换成字形矢量数据,以便输出矢量图形(图 2(e));对笔划作平滑处理,进一步提高笔划的光滑度;视需要作字形变换,以便输出各种变形字。算法过程见图 3。

图 3 中*1: 栅格中心点的坐标值:

$$X = X_0 + d_1(j - 0.5)$$

$$Y = Y_0 - d_2(i - 0.5)$$

式中 d_1, d_2 为栅格的宽和高,

X_0, Y_0 为栅格图原点的坐标,

i, j 为中心栅格的行和列

图 3 中*2: 字形变换公式

$$X = X_0 + HX * KX * \cos(XA) - HY * KY * \sin(YA)$$

$$Y = Y_0 + HX * KX * \sin(XA) + HY * KY * \cos(YA)$$

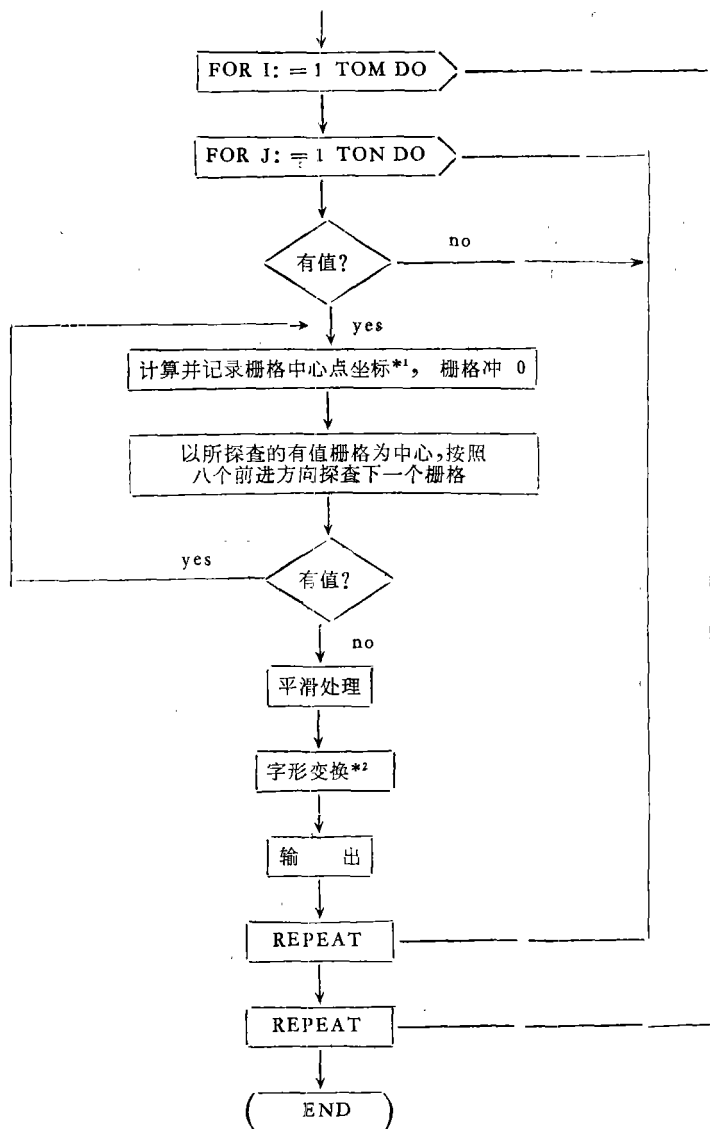


图3 矢量化过程

Fig. 3 The flow chart of vectorization

式中: X_0, Y_0 为汉字左下角点的坐标值,

HX, HY 为字宽与高, 它可控制字的大、小、长、扁等。

$$KX = (i - 24)/48, KY = (24 - j)/48$$

XA, YA 为 X, Y 轴的旋转角度。

当 $XA = YA = 0$ 时为正常字形

当 $XA = YA \neq 0$ 时, 字形未变, 但整个字旋转了一个角度

当 $XA = 0, YA > 90^\circ$ 时为左斜

$YA < 90^\circ$ 时为右斜

当 $YA = 0$, $XA > 0^\circ$ 时为右耸肩

$XA < 0^\circ$ 时为左耸肩

二、软件开发

为了在矢量绘图仪上输出 GB2312-80 国标字库中的汉字、图形符号和英、俄、日、希腊等文字,必须开发一系列的程序。除了实现上述各项功能的程序以外,还有控制整个注

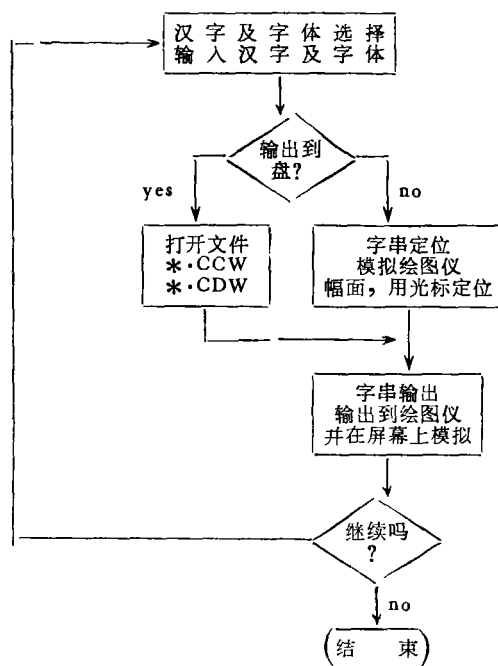


图 4 汉字串输出

Fig. 4 Output of Chinese character strings

记过程的应用程序,例如汉字及其字体选择,注记定位和输出等。注记过程我们设计了三大模块:汉字及其字体选择,定位和输出,如图 4 所示。考虑到图面注记的特点,在程序中各模块均以字串(一系列连续的字符)为单位,而且可使字串注在绘图仪有效范围内任何地方。

选择汉字及其字体,用户可通过区位码、拼音码和首尾码三种方式之一选择所需要的汉字。如果选择的区位码小于 16,则可输出图形符号、数字、外文字等。CCDOS 字库中具备宋、仿宋、黑、楷四种字体,我们将其设计为实心 and 空心两种字形(图 5),前者以大写字母 S、F、H、K 表示,后者以小写字母 s、f、h、k 代表,用户可随意选用。

字串定位,可以用人机对话实时定位,也可以用文件控制快速定位。对于前者,运行 CTT·EXE 程序时,屏幕便开窗模拟绘图仪,按照屏幕提示输入有关参数,按动功能键,即可进行图面注记,见图 6。

功能键的使用,大大方便和加速了字串的定位与输出。

空间数据管理系统
大地海洋科学系

地理信息系统

南京大学

地图学

制图自动化

汉字矢量图形

点阵式字模

正 常 体
右 耸 肩
左 耸 肩
左 斜 体
右 斜 体
长 体
扁 体
※ ※ ※

图 5 字体示例

Fig.5 Some examples of forms of characters

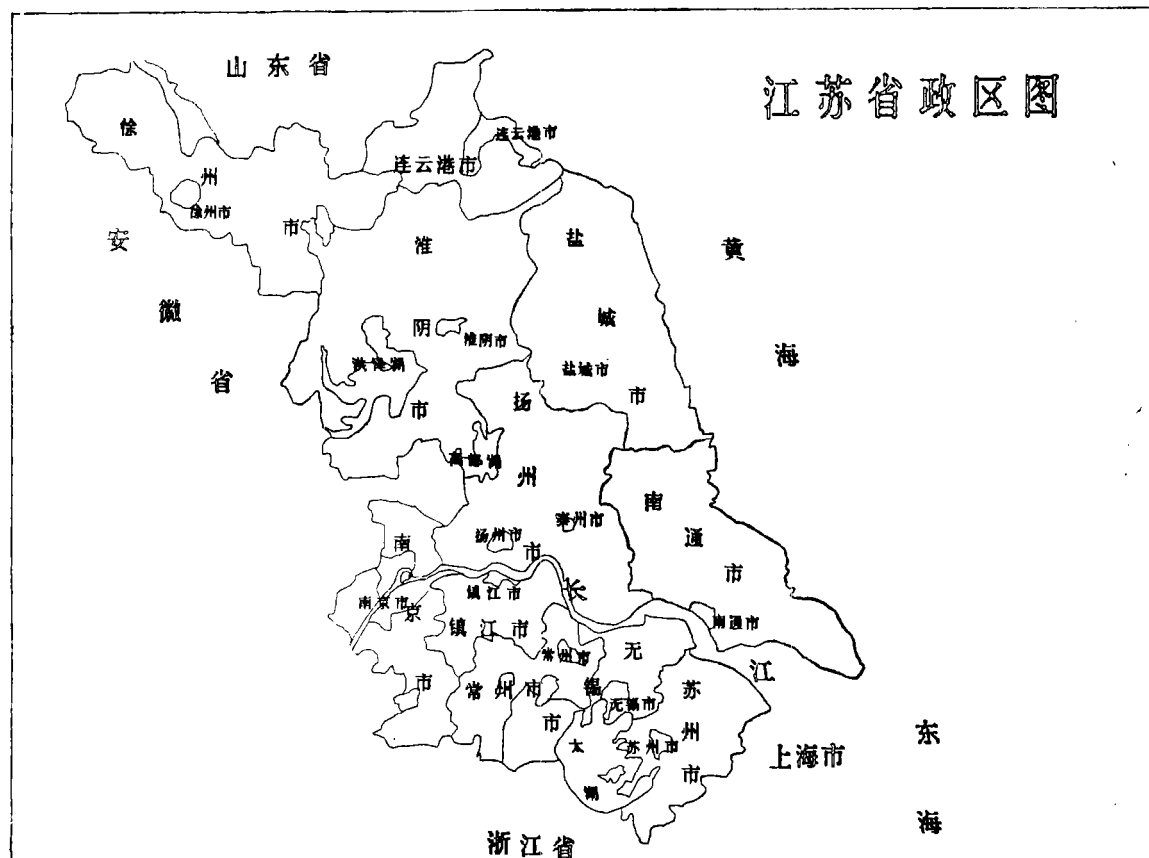


图 6 图面注记示例

Fig. 6 The example of lettering on map

字串不但可以在绘图仪、显示屏和打印机上输出,而且可以输在磁性介质上,形成临时性的汉字注记库 *CCW 及其索引 *CDW,供给不具备汉字处理功能的计算机系统,使它也能进行图面注记。

三、结 论

1. 本文利用和开发国标一、二级汉字信息资源,输出汉字矢量字形,对解决图面注记自动化问题,做了有益的尝试。结果表明,除少数笔划(如撇、捺)有明显的锯齿感,需进一步改进以外,其余都比较光滑,而且字小效果更好。

2. 所开发的软件功能较强,它集汉字显示、打印、绘图和输到磁性介质于一体,以满足用户不同形式的输出要求。屏幕显示与绘图仪绘图同步进行,用光标监控绘图笔,定位灵活、准确。形成临时性汉字注记库,使各种计算机都能进行图面注记。

3. 目前已能写出宋、仿宋、黑、楷四种字体,视需要,可自动扩充任何字体,例如隶体、美术体、自由体等。

4. 从汉字信息处理到矢量字形输出都采用人机对话方式,用户只需运行 CTT·EXE 程序,按屏幕提示输入参数即可。容易掌握和操作。

5. 输出的汉字字形美观与否,同点阵的大小及字库中字形质量有密切关系。随着字库建设的不断改进,输出的矢量字形将取得更好的效果。而且可将矢量字形变成 CCDOS 的组成部分,则图面注记将不受语言的限制。

6. 正式进行图面注记时,必须用刻图针头或光头学绘制,否则影响小字(4毫米尺寸以下)的清晰性。

参 考 文 献

- [1] 张寿萱、徐建设、张建生,中文信息的计算处理,宇航出版社,1984,42—54。
- [2] 唐良招、施国伟,一种新的矢量化方法,微机应用,9(5),1988,19—22。

A NEW APPROACH OF CHINESE CHARACTERS OUTPUT IN VECTOR FORM

SUN YAMEI SUN YIYI

(Department of Geo and Ocean Sciences, Nanjing University)

Key words Chinese characters in vector forms; Automatically lettering on map; chinese Characters Disk Operation System (CCDOS), Matrix of lattices; Raster data of word forms; Vector data of word forms

Abstract

The output of Chinese characters in vector from is the main problem of automatically lettering on map. Chinese characters information process can be done by using user's programs,

operation system or hardware. After a year's research, we found that the first and third usage is not perfect. This paper describes a new approach of using the Chinses characters in CCDOS, and converting matrix of lattices for displaying and printing into vector data for plotting on plotters. It is proved successful by practice that this method can make full use of the cheap Chinese character resource of the first and second levels of character bank in the state standard, and obtain different styles of Chinese characters needed to letter on map.