

AutoCAD. SHX 文件格式剖析、改造及应用

张雅妮 李金钰 陈阳泉

(电子工程系)

摘要 对汉化 Auto CAD(简称 HCAD)中的 SHX 文件进行了分析改造,以 HCAD 中的 MONOTXT. SHX 和 HZTXT. SHX 为基础字库,经过分析、整理和改造,形成字形美观、自由缩放、旋转,易于使用的矢量字体库(包括可视 ASCII 字符集),编写了通用 C 函数,可以满足图形设备(如绘图仪)的汉字输出。

关键词 矢量汉字编码 AutoCAD 形文件 转换程序。

中图分类号 TP311.52 TP391.72

点阵汉字和矢量汉字是两种汉字输出形式。点阵汉字的字形放大、旋转会使字形失真,尤其当放大时,会出现严重的锯齿现象。而矢量汉字将汉字分解成笔划(strokes)矢量集,用画直线的方法来显示这组笔划矢量集,因而字形可以无级变倍仍不失真。建立一种可以为多种软件服务,但不依附于任何特定的图形工具软件的矢量汉字系统是一项十分重要的工作。本文在分析 HCAD 的形文件(SHX 文件)基础上,对其进行整理与改造,已编写出通用的用户接口 C 程序,达到商品化程度。

1 HCAD. SHX 文件格式

HCAD 中的. SHX 文件是由. SHP 文件经过 AutoCAD 编译后形成的。关于. SHX 文件格式在现有文献上没有完整的说明。要以 HCAD 中的. SHX 文件为基础字库完成矢量字体的输出,必须对文件内的数据结构进行正确的提取,其关键在于分析了解. SHX 的格式。

下面以 MONOTXT. SHX(可视 ASCII 字符集)为例分析一般的形文件格式。借助 Advanced PC Tools 5.0 的 File View/Edit Service 功能进行分析,形文件基本结构可以分为两部分:索引表部分及形体(shape)说明部分。

1.1 索引表部分格式

为便于索引表格式描述,作如下约定:“ ”中是 ASCII Value,其它为 Hex Codes,括号中内容是对方括号数据信息的解释及说明。

经分析,索引表格式为如下二进制流:“AutoCAD-86 shapes 1. 0” 0D,0A,1A,00,00[81](结束形代号),00,[64](本形文件中定义的形体个数),00,21,[0A](第一个形体的十六进制代码.由于该文件所描述的是 ASCII 字符集,所以直接以字符代码为索引顺序.HZTXT.SHX 中以汉字机内码代替这部分),00,[08](第一个形体说明字节数),00,⋯,[81](最后一个形体代码),00,[1D](其形体说明字节数),00.

每一个 ASCII 字符的索引为四个字节的,如[81,00,10,00]其中字符代码占两个字节,形体说明字节数占两个字节,每两个字节遵循低位在前,高位在后的原则.

1.2 形体(shape)说明部分格式

“Monospaced TXT Font 11/23/87”00,[06](矢量 ASCII 字符的高度),02,02,00,[00,02,AC,0E,08,09,DA,00](第一个形体说明字节,8个),⋯⋯,[00,02,0E,08,,FF,FA,64,30,01,2A,2C,2E,02,30,0E,3A,0E,28,00](字符“(”的形体说明部分),⋯⋯由此可见,每一个形体的说明部分以“00”开始,以“00”结束.

1.3 HZTXT.SHX 文件结构特点

HCAD 中一般的形文件最多只能定义 256 个形体,而汉字仅一级字库就有三千多个字,矢量汉字形文件(HZTXT.SHX)形成大字体(Bigfont)文件,结构稍有不同.

索引部分,每一个汉字的索引为 8 个字节.其中机内码占两个字节,后两个字节是该汉字形体说明数据的总字节数,剩下四个字节用来表示从文件开始到该汉字矢量数据存贮开始处的字节总数.

如汉字“莫”的索引部分为“AA,C4,61,00,DF,06,03,00”.其中:[AA,C4]为汉字机内码,机内码与区位码的换算公式如下:

$$\text{机内码高位}(C4_H) - 160 = \text{区码}(36)$$

$$\text{机内码低位}(AA_H) - 160 = \text{位码}(10)$$

[61,00]是该字的形体说明字节数,即 61H. [DF,06,03,00]为形体说明部分在 HZTXT.SHX 中的位置:00 03 06DFH=198367.

HZTXT.SHX 中的形体说明部分以“HZTXT-Font 1/81/86”开始.紧接着是 00,[10]汉字的高度),02,00,[00,05,02,54,01,⋯20,00],[00,05,⋯,20,00]⋯每个汉字的说明体以[00,05]为起始标志,以[20,00]为结束标志.

2 HCAD 形文件中的数据代码分析

HCAD 形文件中形体的说明数据是以其专用代码体系来描述的.代码分两种:控制代码与方向代码.

控制代码共 15 种.HZTXT.SHX 中只用到其中的六种.MONOTXT.SHX 中用的较多.为了简化程序设计,对 MONOTXT.SHX 作了修改,使其中的有效控制代码与 HZTXT.SHX 中的相同.六种代码为:

| 代码 | 定义 |
|----|---------------------------|
| 00 | 形定义结束 |
| 01 | 启动绘图方式(落笔) |
| 02 | 退出绘图方式(抬笔) |
| 05 | 将当前位置压入堆栈 |
| 06 | 从堆栈中弹出前位置 |
| 08 | 由下两个字节给出 X,Y 位移,以绘制非标准矢量. |

一般来讲,控制代码 01,02 后紧接一个或几个方向代码.值得注意的是代码 08 后的两个字节表示 X 和 Y 方向位移,实际上是给出了一个点的坐标,其位移为相对坐标. X、Y 位移的范围是-128 到+127.若 08 后两个字节中任一个所表示的数大于 127 即为负数,如 FFH 即为-1,另外,Auto CAD 的坐标系(见图 1)与图形方式下 CRT 的坐标系(见图 2)稍有差别,在编程时应予以重视(绘图仪坐标系同图 1).

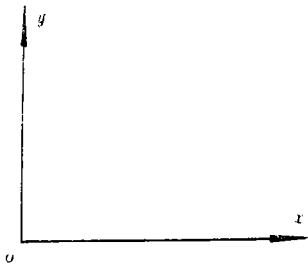


图1 Auto CAD(绘图仪)坐标系

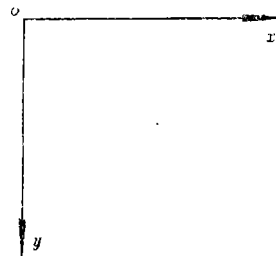


图2 CRT 坐标系

形文件中控制代码高位为 0,低位表示控制信息.方向代码高位表示矢量长度,低位表示方向,整个圆周($0^{\circ}\sim 360^{\circ}$)被分为 16 等份,分别以 0~F 来表示.矢量长度及方向编码见图 3.

对角线上(非水平与垂直方向)的矢量,方向按十六个方向确定,但其大小却是按直角三角形中 X 与 Y 方向矢量中较大者确定的.如“41”表示该矢量在 1 方向,X 位移量为 4,据此来确定矢量长度(见图 4).“4B”表示 B 方向,Y 位移量为 4(见图 5)

从图 3 可以得出:方向 1,7,9,F 的矢量长度都是由 X 方向位移量来确定,Y 方向位移还需再进行计算,在 2,6,A,E 方向,由于 X,Y 位移相等,所以若该矢量长度为 1,则其 X,Y 方向的位移量都为 1,方向 3,5,B,D 矢量长度由 Y 方向位移量确定,X 方向位移还需计算.

下面着重讨论一下计算因子选取问题.设方向代码高位为 K,从图 3 可以得出公式(1),文献[3]中介绍了一种算法见公式(2),编程调试时试过(3)式.

方向 1(22.5°),7(157.5°),9(202.5°),F(337.5°):

$$\begin{cases} |X| = K \\ |Y| = K \operatorname{Tg}(22.5^{\circ}) \text{ (正负另作判断)} \end{cases} \quad (1)$$

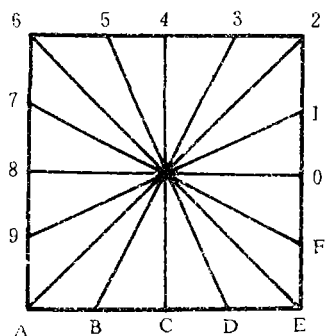


图3 矢量方向编码及长度

$$\begin{cases} |X| = K \\ |Y| = K(1 + \sin\theta - \cos\theta) \end{cases} (\theta \text{ 分别为各方向对应角度}) \quad (2)$$

$$\begin{cases} |X| = K \\ |Y| = K[1 + \sin(1 + \sin\theta - \cos\theta) - \cos(1 + \sin\theta - \cos\theta)] \end{cases} (\theta = 22.5^\circ) \quad (3)$$

相应的方向 3,5,B,D,|Y|=K,|X|计算方法同上.

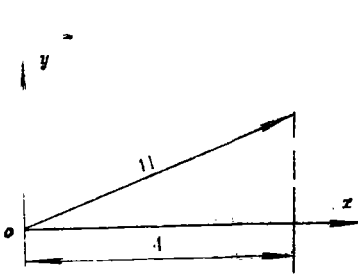


图4 矢量“41”

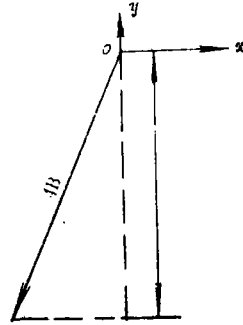


图5 矢量“4B”

分别采用上述三个公式效果都不理想.主要原因是图形方式下CRT 每个像素的坐标都是整数,上述计算结果经取整以后产生误差.

分析三种算法的计算因子:① $\text{tg}(22.5^\circ)=0.412$;② $1+\sin\theta-\cos\theta$.如果 θ 取非 22.5° 的值,例如 67.5° ;则其计算结果为 1.54 ;这显然不符合“长度由X,Y位移较大者决定”的规定.因此,以 22.5° 为 θ 值来计算该因子,值为 0.458 ;③ $1+\sin(1+\sin\theta-\cos\theta)-\cos(1+\sin\theta-\cos\theta)$ 为 0.548 .经过比较,(2),(3)式效果较好.再先取几个中间值试验.最后发现取 0.5 时矢量字体效果最佳(③中 $1+\sin\theta-\cos\theta$ 值单位是弧度).

3 程序编写

C语言具有较强的图形和文件操作功能,控制结构、数据结构丰富.故本程序编写使用Turbo C 2.0.

3.1 文件整理

原HZTXT.SHX索引部分比较杂乱,首先对索引部分按区位码顺序排序以减少查询次数.排序后,原文件索引部分的3200个零字节都集中到文件头部.一般的用户文件既有ASCII字符又有汉字,若在程序中打开两个形文件作为矢量字符库不方便.因此将MONOTXT.SHX(2828字节)合并到HZTXT.SHX零字节部分,这样节省了存贮空间.另外,原MONOTXT.SHX文件输出的ASCII字符不太规范.有些字符不够美观且定位不正确,用PC Tools对其进行了改造.

3.2 索引表数据信息提取

根据索引表部分的结构,分别定义了两个结构体数组.Struct index1表示HZTXT.SHX索引结构,struct index2表MONOTXT.SHX索引结构.

```

struct index1
{
    unsigned int quwei;
    unsigned int zijie;
    unsigned long weizhi;
}id1;

struct index2
{
    unsigned int code;
    unsigned int discribe;
}id2;

```

索引结构体数组中的数都定义成无符号型,以免处理数据时出错,汉字形体说明部分在文件中的绝对位置用四个字节(32位)表示,定义为 unsigned long 型变量。

3.3 形体说明部分数据信息提取

考虑到程序的可移植性及易读性要求,设置两个通用函数分别完成汉字,ASCII 字符的数据信息提取、转换及输出。

(1) chinese (x,y,xtimes,ytimes,quwei)

unsigned int x,y,xtimes,ytimes,quwei;

入口参数依次为:输出时 X,Y 坐标,X,Y 方向放大倍数,该汉字的区位码.依区位码找到相应的 id1.quwei,按 id1.weizhi 定位读取 id1.zijie 个字节,转换输出。

(2) Ascii(X0,Y0,xt,yt,code,start)

unsigned int X0,Y0,xt,yt,code,start ;

入口参数前四个同上.code 表示 ASCII 字符的代码, start 表示第一个字符的形体说明部分在文件中的位置.因 MONOTXT.SHX 索引部分未给出形体说明部分在文件中的具体位置,需用 start 值加上该字符前边所有字符的形体说明字节数来定位。

3.4 辅助程序 VHZM

HZTXT.SHX 共 358KB,占据相当大存储空间.而且每用一个汉字,都要从文件中读取形体说明部分.这样若用软盘,程序运行速度太慢,用硬盘频繁读盘会影响其使用寿命.编写了 VHZM.C 以解决这一问题.程序中,首先对用户文件进行过滤,删除重复的字,然后将文件中汉字的形体说明部分、区码、位码依次赋给三个静态数组,形成 *.C 文件,转换程序中只需用 # include(*.c)就可直接引用有关信息,不用再打开 HZTXT.SHX 来读取数据信息。

3.5 绘图机输出矢量汉字

绘图机与计算机以并行口相接,相当于 C 中的一个基本流 stdprn.用 fprintf(stdprn,“绘图机命令”);就可以把控制信号送给绘图机.不同的绘图机有不同的命令集,程序中只需做相应改动.下面是 SR-6602 型绘图机输出的一个字例样本。

4 结束语

用本文所述的方法建立的矢量汉字系统已在绘图仪(SR-6602)和西文状态下的菜单程序中得到了应用.以本文的工作为基础,还可得到其它专业符号(AutoCAD中已有).进一步工作有以下两点:

- (1)本文所用的 HZTXT.SHX 仅包含一级字库,编写造字程序以满足使用二级字库的要求.
- (2)完成与其他语言的接口.

参 考 文 献

- 1 周克绳,陈德新等编. AUTOCAD 计算机绘图软件(2.6~9.0版).北京:国防工业出版社,1989.
- 2 AUTOCAD 计算机绘图软件(修订本)(2.0~2.6版).北京:中国科学院希望高级电脑技术公司.1987.
- 3 吕勇.如何灵活地在绘图仪上输出汉字.计算机应用研究,1991(1)
- 4 张雅妮.AutoCAD.SHX 文件格式分析及改造应用[学位论文].西安工业学院电子工程系,1993.7

AUTOCAD.SHX file format analysis and modification

Zhang Yani Li Jinyu Chen Yangquan

Abstract

The format of compiled shape file. SHX of Chinese autoCAD(HCAD) is directly analysed and modified. Based on the analysis, re-arrangement and modification of MONOTXT.SHX and HZTXT.SHX in HCAD, an easy-to-use Vectorized Chinese Character Library (visual ASCII set included) is built which can be enlarged and rotated freely and nicely. The functional C function is coded and well tested for more general application purposes. The vector font library of this paper can be applied to output chinese characters in graphic peripherals such as plotters.

Keywords: vectorized chinese character AUTOCAD shape file
converting program