

汉字黑体字形衍生系统的设计与实现^②

岳 华 蔡士杰 顾 进 武港山 严伟荣

(南京大学计算机科学系)

【摘要】计算机在电子印刷、广告、包装等领域应用的迅速扩展迫切需要众多的高质量曲线轮廓汉字库。本文介绍一种通过黑体笔划中心线抽取、笔划轮廓切割、笔划粗细变换、曲线拟合等处理技术来自动生成一族具有黑体风格、但笔划粗细不同的字形的办法。该方法具有低成本、高效率和质量等特点。

关键词: 笔划粗细衍生, 曲线轮廓, 笔划中心线, 汉字库自动生成。

一、引言

不同的应用不仅可能对所用文字的字体有不同的要求、而且可能对字符的笔划粗细有不同的要求。比如, 在印刷业往往使用较细笔划的字符以求得印刷文字的清晰易读; 而在广告业中则往往使用较粗的笔划以求醒目、有力的效果。在条幅标语中, 所用的字符笔划粗细往往随着字符大小而变化, 字符越大, 其笔划应越粗, 才能保证美观。这对字形设计的多样化提出了十分迫切的要求。由于西文字符集小, 使用人工设计方法或计算机辅助的人机交互方法来制备各种字形的字库是可行的。目前西文已有数百、上千种不同的字体供使用。汉字的字符数量非常大, 其一、二级汉字就已近七千个。很难用人工或交互方式来设计数十种乃至数百种不同字体的字库。因此, 寻求用计算机自动生成所需汉字字形的技术势在必行。国内已进行了不少研究, 如文献 (1)、(2) 等作了印刷体生成的实验性探讨, 文献 (3) 对毛笔字体的生成系统作了研究和开发, 国外如日本和韩国 (文献 (4)、(5)、(6)、(7)) 等也对汉字生成与输出描述作了不少研究。本文则介绍作者使用与上述文献不同的方法开发的一个实用的汉字黑体字形衍生系统 CCDS(Chinese Character Deriving System)该系统通过笔划中心线抽取、笔划轮廓切割、笔划粗细变换、曲线拟合、汉字轮廓生成等处理技术, 从一种黑体字形自动衍生出一族具有黑体风格、但笔划粗细不同的字形。实现了低成本、高效率、高质量的汉字黑体字形自动设计。本文将叙述该系统的设计与实现方法。

①本文1994年1月12日收到

②本文得到国家自然科学基金(69273041)资助

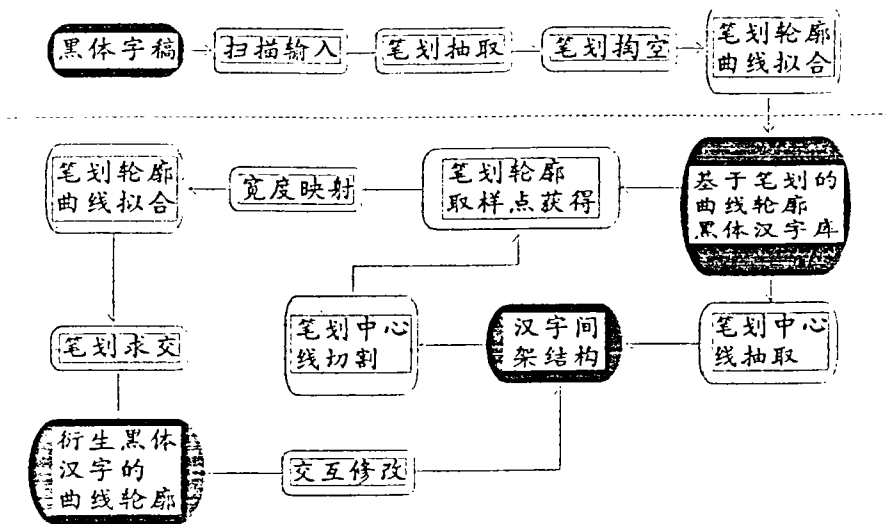


图1 CCDS 的处理框图

二、CCDS 的设计思想及处理框图

汉字黑体字形衍生系统 CCDS 的设计目标是从已有的一套黑体字库自动生成若干套与原字库风格一致的、笔划粗细不同的字库、以支持各种不同的应用场合。因此，以下几点是研制开发 CCDS 的指导思想。

①采用基于笔划的曲线轮廓黑体汉字描述字库、生成完整的曲线轮廓黑体汉字描述，以保证生成字库的高质量。

在汉字字形的各种描述方法中，曲线轮廓描述是比向量轮廓描述与点阵描述均要好得多的一种方法。由于 CCDS 的目标是改变黑体汉字笔划粗细而保持其字体风格，所以其处理的原始数据必须包含笔划信息。笔划粗细的变化不应该改变汉字字形原有的间架结构和笔锋风格，它仅对笔划粗细进行改变，对笔划长度不予改变。

②依赖的初始数据尽量少，降低字形设计成本。

由于汉字字数多，对每个字的衍生处理增加一点数据要求均会对整体增加巨大的工作量，所以 CCDS 除一套基于笔划轮廓的黑体曲线轮廓汉字库外不提出另外的初始数据要求。

③衍生处理应能自动进行，以保证字形设计的高效率。

④CCDS 应提供交互修改工具，用来补救自动处理的某些不足之处。

汉字字形设计实际上是一种艺术造型。用计算机自动处理笔划粗细的改变很难做到百分之百的成功，因此配备一个交互修改工具可以保证质量。

根据上述设计思想，CCDS 的处理流程在图 1 中示出，其中虚线的上部为初始数据制备流程，本文对此不作介绍。

三、基于笔划的黑体汉字曲线轮廓表示及其间架结构的抽取

3.1 黑体汉字笔划的曲线轮廓表示

CCDS 的核心问题是实现对黑体汉字笔划粗细的控制。因此,它处理的数据必须有笔划信息。CCDS 使用的源数据是一个基于笔划曲线轮廓的黑体汉字库。它是从黑体汉字字稿经过扫描输入转换成点阵信息,再经过笔划抽取、笔划掏空、轮廓曲线拟合等工作而得。

在该汉字库中,每一汉字(称为源汉字)由若干个笔划组成,每个笔划由一个按逆时针方向走的封闭曲线环描述(图 2)。这样的源汉字可使用 BNF 形式描述如下:

源汉字 ::= <笔划> {<笔划>}

笔划 ::= <曲线段> {<曲线段>} (末段的终点与首段的始点重合)

曲线段 ::= <折线段> | <三次 Bezier 曲线段>

折线段 ::= <始点> {<点>} <终点>

三次 Bezier 曲线 ::= <始点> <点> <点> <终点>

始点 ::= <点>

终点 ::= <点>

点 ::= <x> <y>

x ::= 整数

y ::= 整数

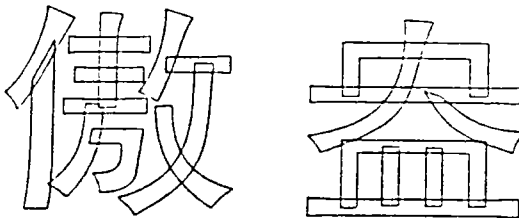


图 2 源汉字的描述

3.2 笔划路径的判定

CCDS 只改变笔划的粗细而不改变其长度,因此必须识别笔划的路径,即找出其始端、终端及两侧边。然后对两侧边的相互距离按所需笔划宽度加以控制修改。

笔划中的端段有以下几个特征:

- 段内曲线是光滑的。
- 长度与笔划粗细接近。
- 该曲线段两端为奇点。
- 与其两端相接的两曲线段为两个侧边,它们与端段之间的角度差较大,且均在该段的同一侧(至少其初始部分是如此)。

笔划的侧边有以下特征:

- 除“横折”等复合笔划以外，多数笔划的侧边上无奇点。
- 长度一般比笔划宽度大得多。
- 两侧边的距离与笔划宽度接近。
- 在两侧边上可找出若干成对的互相（近似）平行的小段。它们构成两侧边的主要部分（图3）。

依据上述两组特征，分析一笔划的轮廓，可以识别侧边与端段。

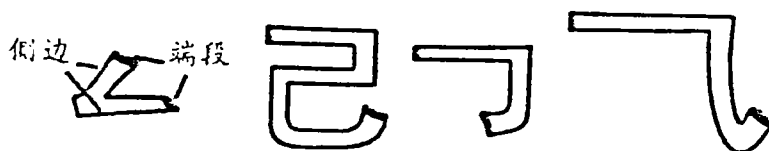


图3 笔划的侧边与端段

3.3 笔划中心线的抽取

严格地讲，笔划中心线是与两侧边等距离的曲线。抽取笔划中心线的基本方法是找出两侧边上若干组对应点，求出它们的连线中点，然后将这些中点排序后拟合成一条或多条曲线。由于侧边并非均为直线段，故寻找对应点的难度较大。我们使用将侧边分段移动并进行连接处理的方法来求得笔划中心线。

① 侧边直线段的移动

设该笔划宽度为 W ，将该直线段向轮廓线行进方向的左方平移 $W/2$ （图4a）。

② 侧边的三次 Bezier 曲线的移动

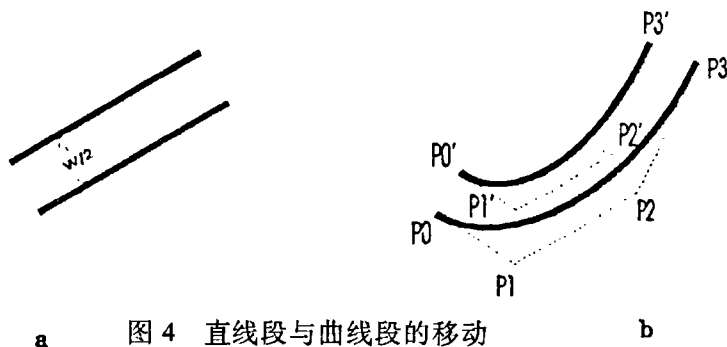
设该 Bezier 曲线的四个控制点分别为 P_0 、 P_1 、 P_2 和 P_3 ，分别将直线段 P_0P_1 、 P_1P_2 和 P_2P_3 按①中规则移动，获得 P'_0 、 P'_1 、 P'_2 、 P'_3 为中心线上对应 Bezier 曲线段的控制点，其中 P'_1 为 P_0P_1 和 P_1P_2 移动后的交点， P'_2 为 P_1P_2 和 P_2P_3 移动后的交点。（图4b）。

③ 原来为平滑连接的两段，在移动后保持为平滑连接，且公共端点移动后仍公共。

④ 原来非平滑连接的两段，在移动后需经求交运算获得新的公共端点。

⑤ 连接处理后的中心线，其两端经与两端段求交而得。

图5给出了源汉字及抽取出的笔划中心线的例子，这些中心线构成了该汉字的间架结构。



a

图4 直线段与曲线段的移动

b

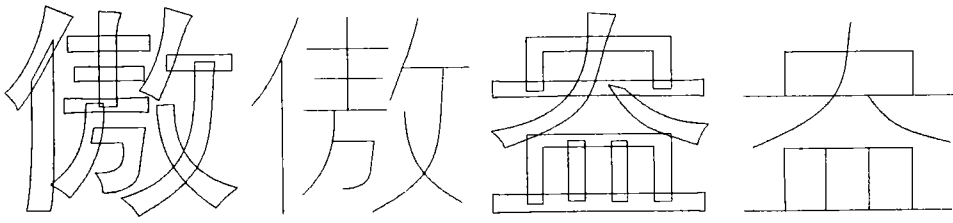


图 5 源汉字与间架结构

4. 笔划的衍生及衍生汉字字形的生成

4.1 笔划衍生原理

笔划的衍生是根据原有的笔划在保持类型与风格的前提下, 改变其粗细 (即笔划宽度) 生成新的笔划。一般情况下, 原笔划轮廓上的一点 P 与其衍生笔划上对应点 P' 有如下关系:

$$P' = O + s * (P - O)$$

其中点 O 是过点 P 、 P' 的直线与笔划中心线的交点, P' 、 O 、 P 分别为三个点向量。 s 为衍生宽度比。

4.2 取样点和衍生点

上述点 P 称为原笔划轮廓取样点, 点 P' 称为点 P 的基于宽度因子 s 的衍生点。取样点通过笔划切割来实现, 笔划切割首先从笔划中心线开始。设 $O_i(x_{oi}, y_{oi})$ 为将中心线等分 n 段的第 i 分点, 作 $O_i P_{i1}$ 与中心线在 O_i 处垂直, 交笔划轮廓于 $P_{i1}(X_{pi1}, Y_{pi1})$ 和 $P_{i2}(x_{pi2}, y_{pi2})$, P_{i1} 、 P_{i2} 即为对应于 O_i 的两个取样点。此时, 点 $P'_{i1}(X'_{pi1}, Y'_{pi1})$ 和 $P'_{i2}(X'_{pi2}, Y'_{pi2})$ 为 P_{i1} 、 P_{i2} 的 s 衍生点, 只要

$$X'_{pi1} = X_{oi} + s * (X_{pi1} - X_{oi})$$

$$Y'_{pi1} = Y_{oi} + s * (Y_{pi1} - Y_{oi})$$

$$X'_{pi2} = X_{oi} + s * (X_{pi2} - X_{oi})$$

$$Y'_{pi2} = Y_{oi} + s * (Y_{pi2} - Y_{oi})$$

4.3 取样点与衍生点的特殊处理

对于笔划中心线为多段且不平滑连接的情况, 中心线曲率半径较小的部分以及笔锋处, 必须对取样点或拟合点作一些特殊处理。

如图 6 所示, 笔划中心线由两段曲线组成, 连接处并不平滑。其内侧轮廓上取样点

An(1)应是对应于前一段中心线的最后一内侧取样点，而在外侧轮廓上 BW(1)与 BW(2)之间无正常取样点。此时可作一组射线来获取一组取样点，使 Bw(1)和 Bw(2)之间有足够量的取样点去生成对应衍生点。

笔划中心线不平滑时，转折点被用来对内侧取样点分组，两组取样点所得衍生点分别去拟合成两条曲线。它们要进行求交运算后才能获得衍生后的内侧轮廓。笔划中心线光滑但曲率半径较小时，加粗笔划时的内侧衍生点也要进行处理，如图 7 所示，A、B、C、D 分别为按先后次序获得的衍生点。由于 B 的位置已位于 A 之前，构成衍生点的“回头”情况，如保留，必定使拟合曲线质最降低，故必须舍弃 B 点。基于对称性构成衍生点“超前”的 C 点也应被舍弃。因此，在处理这样的笔划部分时，要先对弯曲程度作一扫描，找出弯度最大处，以此为中心从两边向中心方向作衍射点的弃留处理。

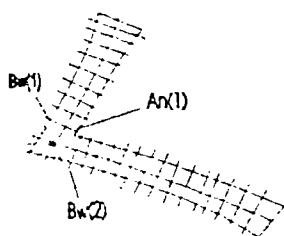


图 6

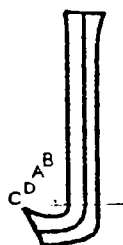


图 7

4.4 笔锋处理

在图 8a 所示的笔划中，使用正常的方法得不到笔锋部分的取样点，如图 8a 中对应于中心线末点的取样点，离内侧边末点还有一段无取样点，因此也无法获得衍生点。如果让已有衍生部分和笔划端曲线延长求交来获得笔锋，则可能会造成较大的变形（图 8b）。我们通过以中心线末端为中心对该部分幅射比例变换的方法来获得衍生点（图 8c），从而保持了笔锋特征（图 8d）。

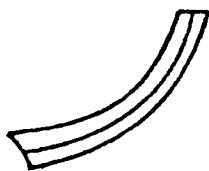


图 8a

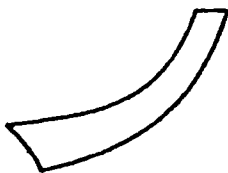


图 8b

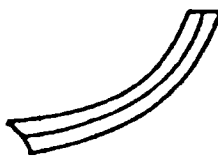


图 8c

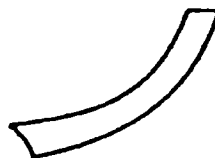


图 8d

4.5 轮廓拟合与组织

经过衍生而得的笔划应与原笔划有同样的表示，即逆时针曲线轮廓表示。因此，在获得所有的衍生点后还要进行下列处理：

①将每一组衍生点拟合出一 Bezier 曲线来，并且和源汉字相应笔划比较，保证和源

汉字曲线方向一致。

②将每一侧边对应的衍生拟合曲线连接起来，并修改端边，必要时进行求交运算。

③将其中一顺时针的新侧边的首尾互换，形成逆时针侧边。

④连接成封闭的笔划轮廓线。

4.6 交互修改

由于汉字的结构十分复杂，很难通过抽取若干特征而覆盖所有汉字笔划。在原始的黑体数据中，存在一些连笔划的问题，如图 9 所示。另外，对于笔划数较多的黑体汉字，当宽度因子增加到一定的值时，会造成笔划重叠的现象。对这些现象本系统未专门处理，留待交互修改。所以衍生而得的基于曲线轮廓笔划的汉字中难免有个别地方不尽如意。因此提供一个交互修改程序是有益的。该交互程序的主要功能为取 / 存字；选择笔划；选择顶点 / 控制点；移动顶点 / 控制点；直线段和 Bezier 曲线的互相转换；删 / 增曲线段等。在交互修改时，以比较淡的不同颜色显示出原字形以助于正确修改。



图 9

4.7 笔划间求交

最终提供使用的衍生汉字，应是完整的曲线轮廓表示。即必须将有交叉的笔划经求交运算以取得外轮廓。

求交工作分两步进行：①求得交点。②修改曲线描述。

笔划间交点的获得通过指定精度下的 Bezier 曲线二分法来实现，由于主要是加减与移位运算，故速度较快。交点所在的 Bezier 曲线必须一分为二，其中一段保留在最终曲线轮廓中，另一段则丢弃。

如果由控制点 Q_0 、 Q_1 、 Q_2 和 Q_3 确定的一条 Bezier 曲线上交点的参数为 t_0 ($0 < t_0 < 1$)，则该点前后两段 Bezier 曲线可分别表示为：

$$B_1(u) = [u^3 \ u^2 \ u^1 \ 1] M_B \begin{bmatrix} Q_0 \\ (1-t_0)Q_0 + t_0Q_1 \\ (1-t_0)^2Q_0 + 2(1-t_0)t_0Q_1 + t_0^2Q_2 \\ (1-t_0)^3Q_0 + 3(1-t_0)^2t_0Q_1 + 3(1-t_0)t_0^2Q_2 + t_0^3Q_3 \end{bmatrix}$$

$$= [u^3 \ u^2 \ u^1 \ 1] M_B \begin{bmatrix} Q'_0 \\ Q'_1 \\ Q'_2 \\ Q'_3 \end{bmatrix} \quad (0 \leq u \leq 1)$$

$B_2(u)$

$$= [u^3 \ u^2 \ u^1 \ 1] M_B \begin{bmatrix} (1-t_2)^3 Q_0 + 3(1-t_0)^2 t_0 Q_1 + 3(1-t_0)t_0^2 Q_2 + t_0^3 Q_3 \\ (1-t_0)^2 Q_1 + 2(1-t_0)t_0 Q_2 + t_0 Q_3 \\ (1-t_0)Q_2 + t_0 Q_3 \\ Q_3 \end{bmatrix}$$

$$= [u^3 \ u^2 \ u^1 \ 1] M_B \begin{bmatrix} Q''_0 \\ Q''_1 \\ Q''_2 \\ Q''_3 \end{bmatrix} \quad (0 \leq u \leq 1)$$

其中

$$M_B = \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 3 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

五. 总结

本文所述的方法已经在 386/SX33 上用 C 语言实现。程序量约为两万行语句。在有 80387 协处理器的情况下，生成一套指定笔划宽度的黑体字字库（一、二级汉字）需运行约 80 小时。其中约 10% 的字中个别笔划要作一些修改，修改时间为 4 天。生成一个汉字所花费的时间不仅依赖于汉字的笔划数，而且还依赖于该汉字中弯曲程度较大的笔划的数目。表 1 中给出了八组汉字的生成时间对比信息。如第①、②、③组汉字，上面的字虽然笔划数比下面的字少，但由于有弯曲笔划，因而生成时间长。又如第④、⑤、⑥组汉字，上下两个字的笔划数相同，汉字中弯曲笔划多的字生成时间长。附录中给出了衍生而得的不同粗细的黑体字的例子，由此说明了本文所述的方法生成汉字字形的效率很高，效果也好。

表 1 (时间以秒为单位)

八 30.48	必 39.17	包 28.84	艾 43.83	大 29.93	捺 64.70	笆 34.71	办 30.70
蠶 14.44	鼻 20.37	蚌 24.27	半 20.22	寸 12.64	曹 18.35	吧 15.76	絆 37.29

参考文献

- [1] 樊建平“智能汉字字形设计技术及一个试验性系统 ICCDS”, 中文信息学报, 1990年 Vol.4, No.3
- [2] 樊建平“基于汉字结构码量化传统书法规则知识方法的实现”, 中文信息学报, 1990年 Vol.4, No.4
- [3] Xhang Xianrong, “High Precision Small–Volume Beush–Written Chinese Characters with the Pattern Synthesis of Hierachical Synthesis Analysis” Proceedings, 1990 International Conference on Computer Processing of Chinese and Oriental Languages, April 25–28, Changsha, China, PP 266–271
- [4] Zengo FURUKAWA, Togomi KAWAMURA and Kazuo USH IMA, “Evaluation of Japanese Character Fonts on a PostScript Printer” Proceedings, 1990 International Conference on Computer Processing of Chinese and Oriental Languages, April 25–28, Changsha, China, PP255–260.
- [5] S.B. Lim, Outline Font Representation with Structured Description of Characters, Ph'D Dissertation, Department of Computer Science, KAIST, 1992
- [6] H.Chen and S. Ozawa, “A Methodical Generation of Various Kinds of Mincho Charater”
日本情报通信学会论文志 D–H, Vol. J72–D–11, No.9, PP.1423–1431, 1989
- [7] H.Ergong etal, “Multi–Demension Parameter Contral Technique”. Intern'l conf. on Computer Processing of Chinese and Oriental Languages, PP182–185, 1988.

The Design and Implementation of a Chinese Character Deriving System for Heiti

Yue Hua Cai Shij ie Gu Jin
Wu Gangshan Yan Weirong

(Computer Science Department, Nanjing University, 210008)

Abstract

The rapid extension of computer application in the areas of electrical publishing, advertising and decoration design raises the needs for variant chinese character font with high quality curve–outling representation. This paper introduces an automatic chinese font generation method for Heiti using the techniques of stroke center–line extraction, stroke outline deviding, stroke thickness variation and curve fitting. The method has the advantages of low cost, high efficiency and quality.

Key Word: Stroke chickness deriving, curve outline, stroke centerline, automatic chinese font generation.

