

一种从黑体到隶书的汉字字形自动变体方法*

高 晓 蔡士杰

(南京大学计算机科学与技术系, 南京 210093)

摘要 本文介绍了一种以构架和基本笔划为汉字的知识表示, 从黑体到隶书的汉字字形自动变体方法. 该方法首先实现基于特征的黑体构架到隶书构架的变换, 然后实现基于笔划柔性映射的隶书构架到隶书轮廓字形的生成. 此方法探讨了一条汉字字体间自动转换的途径, 并具体实现了从黑体到隶书的自动变体.

关键词 变体, 黑体, 隶书, 构架, 柔性映射.

电子出版业在我国的迅猛发展和计算机在广告行业等方面的应用, 迫切需要品种繁多的汉字曲线轮廓字体. 但由于汉字属大字符集, 且结构复杂, 要用手工方法设计出多套曲线轮廓汉字字体, 在时间和成本上都是难以承受的. 因此, 借助于计算机, 从一套基本字体出发, 通过变体技术, 设计出多种汉字字体以满足各领域应用的需求, 具有十分重要的意义. 在这方面, 作者进行了一些探索、研究与开发. 文献[1]给出了从一种黑体字体自动衍生出多种笔划粗细不同的黑体字体来以便适用于不同的应用场合. 文献[2]则给出了以黑体构架为基础自动生成多种粗细规格的圆头体和等线体字体的方法. 加速了计算机汉字字体的开发.

本文介绍汉字从黑体自动变换成隶书的一种方法. 该方法根据汉字字形设计的两大要素: 构架和笔划, 首先实现从黑体构架到隶书构架的转换, 然后再实现笔划的映射, 从而实现字体的自动转换.

1 两种字体的异同点分析

黑体是一种常见的印刷字体. 它字形方正、横平竖直、起落笔整齐、笔端统一、笔划变化少、粗细均匀.

隶书是篆书的变体, 楷书的前身, 是我国传统的艺术书法之一. 它字形扁阔、体势开张、向背分明、整齐舒展. 笔划雄厚矫健, 笔法富有变化. 庄重大方、气淳质朴.

黑体和隶书有两点很重要的相同之处. 一是黑体和隶书具有相同的拓朴特征. 二是黑体和隶书都有清楚、明确的笔划划分, 且黑体笔划所属的类与它在隶书中所属的笔划类是一致的.

* 本文 1994-04-01 收到, 1994-06-27 定稿

本研究得到国家自然科学基金的资助. 作者高晓, 女, 1969年生, 1994年硕士毕业于南京大学, 主要研究领域为计算机图形学与CAD. 蔡士杰, 1944年生, 教授, 主要研究领域为计算机图形学, CAD, 用户接口.

本文通讯联系人, 蔡士杰, 南京 210093, 南京大学计算机科学与技术系

但黑体和隶书毕竟是两种不同的字体,有其各自的书法特征.具体来说,黑体和隶书字体的差异主要表现在以下 5 个方面:

(1) 从整体字形上看:黑体字形方正,呈正方形;隶书字形方正而扁平,呈扁长方形.

(2) 从笔划种类来看:黑体笔划统一,笔划种类少;隶书笔划多变化之美,笔划种类多.

(3) 从一个字中各笔划的粗细变化上看:黑体在同一格内,不受笔划繁简限制,一般要求笔划一致,整齐化一;在隶书中,主要笔划墨迹浓重有力,次要笔划墨迹轻清,粗细相间,疏密得宜.

(4) 由于隶书是由篆书演变而来,它的偏旁部首的写法,很多地方和黑体不一样.

(5) 黑体是一种印刷字体,规则约束性强,字体统一;隶书是一门书法艺术,无完全严格的统一规则可循,字体多变化.

黑体和隶书两种字体的相同之处是进行黑体到隶书变体的基础.两者拓朴结构上的一致性,使得从黑体构架到隶书构架的转换成为可能.两者笔划分类的总体一致性,使得笔划映射得以进行.而黑体和隶书两者字体上的差异,则是进行变体的依据.只有在从黑体到隶书变体过程中更多地反映这些差异,才能更贴切、生动地体现隶书的书法特征,使变体成功.

2 变体方法的设计

2.1 基于构架和笔划柔性映射的汉字字形表示方法

汉字的字形随字体的不同而呈不同的书法风格,这是由汉字的笔划特征、几何特征和拓朴特征所决定的.汉字的几何特征和拓朴特征决定了汉字的构架,构架不同,汉字字形也就发生了变化.笔划特征通过一些基本笔划来反映,可以把基本笔划组成基本笔划库,它与构架一起来确定字形.

本文使用的就是一种基于构架和笔划柔性映射的汉字字形知识表示方法,就是
 构架库 + 基本笔划库 = 一套汉字库.

(1) 构架库

汉字的构架信息用以给出该汉字的组成部件、各部件的组成笔划等拓朴信息,以及每一笔划的路径等几何信息.这就是说,一个汉字的构架信息的数据结构是层次式的.下面是一个汉字构架的 BNF 描述:

〈构架〉 ::= 〈部件数〉 { 〈部件〉 }

〈部件〉 ::= 〈部件类型〉 〈部件属性〉 〈内含笔划数〉 { 〈笔划〉 }

〈部件属性〉 ::= 〈部件位置〉 〈部件大小〉

〈笔划〉 ::= 〈笔划类型〉 { 〈段描述〉 }

〈段描述〉 ::= 〈段几何类型〉 〈段几何数据〉

〈段几何类型〉 ::= 〈直线段〉 | 〈三次 Bezier 曲线段〉

每一汉字构架中部件的划分、笔划的划分、笔划路径的方向以及笔划的顺序等均按习惯的书写规则进行.其中,部件的位置是指它在该汉字横向结构的“左、右”、“左、中、右”部,或纵向结构的“上、下”、“上、中、下”部,或其它的“两面包围”、“三面包围”、“全包围”以及“三叠四叠”结构中的某一部分.这对其组成笔划的形状和粗细有一定的影响.部件的大小是指该部件在所在汉字中的某种结构中相对于其它部分的大小,分适中、偏大、偏小 3 种情况,这对

其组成笔划的长短和粗细有一定的影响. 同样, 部件内所含的笔划数也会影响部件内各笔划的粗细. 所以, 结合部件的属性和字体的特征调整部件内所包含笔划的长短、粗细、形状等, 能更精确地对字形进行描述.

(2) 基本笔划库及笔划的柔性映射

每一种字体都可以分解并提取出一组基本笔划, 它们和汉字的构架一起在汉字的书写中反映出这种字体的特征. 一种字体的基本笔划分类数目与笔划形状的提供之间存在着矛盾. 笔划分类太多会使笔划库的规模过大, 笔划分类若太少又会使笔划映射所生成的汉字字形的质量受到影响, 不能很好地体现字体所具有的特征与风格. 事实上, 有些笔划只是在长短、粗细、或行笔路径上有差异, 而它们的曲线轮廓的组织是完全相同的, 可以将这些笔划归为一种基本笔划.

基本笔划库的组织和使用既要保证它能提供所需的足够多的笔划品种, 以满足描述某种字体、体现其独特风格的要求, 又要使品种数量即库的规模尽量地小. 下面是笔划库中每一基本笔划的 BNF 描述:

```

<笔划> ::= <始端描述> <终端描述> <左侧边描述> <右侧边描述> <中心线描述>
<描述> ::= { <直线段> | <三次 Bezier 曲线段> } <描述结束标志>

```

每一基本笔划的信息中给出了笔划的轮廓数据和中心线数据以完整描述笔划的形状及行笔路径. 其中, 始端描述和终端描述描述的是笔划的笔锋. 按逆时针方向把始端描述、左侧边描述、终端描述、右侧边描述顺序组织, 即是对笔划外轮廓形状的描述. 而中心线描述则给出了笔划的行笔路径.

以基于构架和基本笔划库为字形表示方法, 具体构作某一字体的汉字字形时, 必须灵活运用基本笔划, 即将基本笔划和汉字的构架结合起来. 也就是说, 同一种基本笔划用在不同的汉字中, 或同一汉字的不同位置, 其最终的几何特征, 如长短、粗细、行笔路径不一定完全相同, 这就需对笔划进行柔性映射, 即根据汉字构架中这一笔划的几何特征, 对基本笔划库中相应笔划的数据进行适当的变换、处理, 再最终映射到具体汉字的具体笔划上.

笔划柔性映射的主要处理思想是: 保证映射后笔划侧边线上的点到其中心线的距离与基本笔划上对应点到中心线距离之比等于宽度因子. 而两中心线形状可以不同. 图 1 为同一基本笔划的两种不同映射.



图 1

其中, 宽度因子是为反映基本笔划库中笔划的长短、粗细与在构作汉字时具体实例的差异设置的. 一般地, 宽度因子可取汉字构架中笔划的路径长度与基本笔划库中相应笔划的中心线长度的比值.

2.2 从黑体到隶书的汉字字形变体处理

由于黑体和隶书具有基本相同的拓朴结构, 故在黑体构架的基础上, 融入隶书的几何特征, 可变换生成隶书的构架; 再以隶书的构架为基础, 结合反映隶书书法风格的基本笔划库, 通过笔划的柔性映射, 就可生成隶书的轮廓字形, 从而实现了从黑体到隶书的变体. 图 2 是

整个变体过程的总体流程图.

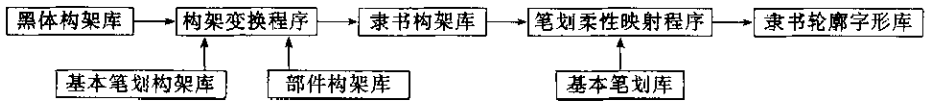


图2

3 基于特征的构架变换

从黑体构架到隶书构架的变换,是整个变体过程的关键之一.变换过程中所用到的数据有:一套黑体构架库和一套隶书基本笔划构架库.其中,黑体构架库包括对每个黑体汉字的构架描述.隶书基本笔划构架库则是抽取隶书的每一基本笔划在基本笔划库中的中心线描述信息,以及相对应的黑体笔划的中心线信息所构成的.下面详细叙述构架变换的一般规则及对各特征差异的处理.

3.1 一般规则

从黑体到隶书构架变换的一般处理过程是:(1)预处理:对组成黑体构架的每一笔划,根据其笔划类型,从基本笔划构架库中取出相应笔划的黑体和隶书中心线数据进行处理,使隶书笔划在保持其基本风格不变的前提下与黑体笔划有最相近的闭包框(比例变换).(2)替换:首先比较黑体构架中将被替换的这一笔划和相应库中的黑体笔划,求出它们闭包框的 x 向比例因子和 y 向比例因子,并把这一比例因子作用到库中已经规范化了的隶书笔划上(比例变换).然后,将经过上述处理的隶书笔划平移到黑体构架中要被替代的笔划处(平移变换),使两者重心相重合,并取代相应的黑体笔划.

3.2 部件替换

从黑体构架到隶书构架的变换,一般是以笔划为基本处理单位进行的.但注意到,由笔划组成的部件一方面是构成汉字的不宜分拆的基本结构单位;另一方面,由于一些部件在隶书中的写法与在黑体中的不尽相同,如笔顺甚至拓朴结构上的差异,如图3例,使组成这类部件的笔划,必须以部件为单位进行替换处理.由于部件中浓缩了隶书的许多书法特征,所以以部件为单位替换能更集中、突出地反映出隶书和黑体字形上的特征差异,同时也能提高处理速度和效率.



(a)黑体和隶书部件在拓朴结构上的差异 (b)黑体和隶书部件在笔顺上的差异

图3

3.3 重心的选择和调整

在笔划替换或部件替换中,平移的重心一般选取的是笔划闭包框(笔划替换)或部件闭包框即组成部件各笔划的闭包框(部件替换)的几何重心,这是最常用也最具普遍意义的作法.但在某些情况下,虽然用以替换的隶书笔划或部件与黑体构架中被替换的笔划或部件这

种意义上的重心是一致的,但如此替换后会导致整个字的结构不匀称,故需做调整.

例如,对笔划“竖钩”替换时,若仍按通常的闭包框几何重心重叠的方法,如图 4 所示,会造成整个字布白不均、重心不稳.为解决这一问题,对这类笔划,可选择其主笔划段(即质量集中的笔划段),如“竖钩”中起笔的“竖”段,为平移时重心选择的依据.



(a) 黑体笔划与重心 (b) 隶书笔划与重心 (c) 闭包框几何重心重叠法 (d) 几何重心替换法 (e) 主笔划段替换法

图 4

在部件替换时,有时需选择部件中的主笔划而非部件闭包框的几何重心作为其平时重心的依据.如图 5 所示,选择耳朵旁中的笔划“竖”为其主笔划.



(a) 黑体中的耳朵旁 (b) 隶书中的耳朵旁 (c) 选“竖”为其主笔划使平移后重叠

图 5

另外,有些部件,它在隶书中与在黑体中所处的纵向位置本身就有明显差异,需予以适当调整后才能准确地反映出隶书的特征.如部件“口”,当它作左偏旁时,在黑体中一般放在一个字的“腰”部,即部件“口”闭包框 y 方向的几何重心落在整个字高度的 1/2 处;而在隶书中,一般多在整个字的上半部,即部件“口”闭包框的 y 方向几何重心落在字高的 1/4 处上下.如图 6 所示的字例.



黑 体 隶 书

图 6

3.4 整体比例调整

隶书闭包框的长宽比一般取为 10:7,这只需对变换后得到的隶书构架的 y 方向在整体上做一比例变换即可.

3.5 笔划类型的对应

黑体笔划所属的类与映射后的隶书笔划所属的类在总体上是一致的.但由于黑体的笔划较单一,笔划种类少;而隶书的笔划多变化,笔划种类较多,所以黑体笔划类型与映射后的隶书笔划类型之间多为一多对应关系.映射后的隶书笔划类型的确定,需根据此笔划在其所在部件中和部件内其它笔划的笔划类型间的相互关系,以及此笔划在部件中甚至整个字中所处的位置等情况具体分析而定.

例如,黑体中的笔划“横”映射到隶书中时,它一定仍属于隶书的“横”笔划类,这就是黑体和隶书映射前后笔划类的总体一致性.但隶书中的“横划”又可分为“波划”与“平划”两类,其中“平划”根据其笔势又可细分为“仰势”与“覆势”两种,如图 7 所示.于是,黑体中的笔划类型“横”与隶书相应的笔划类型之间存在着—多对应关系.

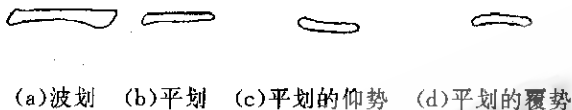


图 7

对于“波划”,一般称其头部的笔锋为“蚕头”,称其尾端的捺脚为“燕尾”.在隶书中有两句口诀:“燕不双飞”、“蚕无二设”.其中“燕不双飞”的意思是说,一个字中捺脚要避免重复(包括“波划”中的捺脚和笔划“捺”中的捺脚),即一个字中最多只能有一个“燕尾”.这就是说,只有当“横划”的尾端是整个字中所有笔划的最右端时,才有可能具有“燕尾”.“蚕无二设”的含义是,在一个部件中,当连续有两个或两个以上的笔划是“横划”时,其中不可有两个以上的“蚕头”出现.同样,在这种“横划”并列的情况下,“平划”有仰、覆势之别.一般取较上面的一笔为覆势,而取较下面的一笔为仰势,这样整个字的各笔划之间才有呼应、有照顾.

3.6 两类特殊笔划的处理

对“撇”和“捺”这两类特殊笔划,由于它们的行笔路径变化多种,很难以基本笔划构架库中的单一模板经过简单的平移、比例变换后去替代各汉字构架中的具体实例,需采用称为“导向技术”的特殊处理.

具体的处理方法是:首先对汉字黑体构架中要被替换的这一笔划的笔划路径(记为①)以及相应基本笔划构架库中这一笔划的黑体中心线(记为②)和隶书中心线(记为③)三者做一定的平移—旋转—比例变换,使得变换后这三者的首、尾端点均重合,如图 8 所示.由于所做的均为线性变换,所以三者变换后的几何形状与原先的是相似的.其次,用一组短



图 8

直线段对②均匀地正交切割,分别对每一根切割线,取其于②和①交点间的直线段长度,记为 l .然后,用另一组数目相同的短直线段以同样的均匀度对③进行正交切割,对每一根切割线,取其于③交点同侧(所谓“同侧”的含义是:若①在②的右侧边,则取交点右边的线段上的点;反之亦然). l' 距离的点.在这里,若取 $l'=l$,则称为完全导向;若 $l'=0$,即取库中原隶书的中心线为替换后隶书构架中相应笔划的路径,称为完全不导向;若 $0 < l' < l$,称为部分导向.称 l' 与 l 的比为导向因子.导向技术即是通过导向因子来控制两类特殊笔划在最终生成的隶书构架中的笔划路径的.最后,对切割所生成的一组点进行拟合,即可生成最终用以替换黑体构架中相应笔划的隶书笔划路径.

3.7 特色笔划组的替换处理

隶书中的某些笔划,当它们和其所在部件中的其它笔划一起形成特定的笔划组合时,它们各自的几何形状较它们单独存在时有一些改变,它们相互间的拓朴关系较它们存在于黑

体构架中时也会有所变化. 对此, 必须进行适当的调整, 才能反映出隶书和黑体字形上的差异, 更准确地体现出隶书的风格. 称这些具体地体现了隶书特色的笔划组合为特色笔划组, 对它们的处理方法如同部件替换类似. 下面举两个特色笔划组的例子加以说明.

(1) 特色笔划组“口”

隶书“口”和黑体“口”不仅在几何特征上存在着差异, 它们的拓朴结构也有所不同(如图9所示): 这三个笔划组合的路径在隶书中是封闭的而在黑体中则不是. 为此, 对这三个笔划分别进行替换不能反映出隶书与黑体的这一差异, 必取提取这三个笔划的组合, 把它们作为一个整体, 即特色笔划组来加以处理.



图9

(2) 特色笔划组“冂”

在某些情况下, 黑体中的“竖”可能会映射成隶书中的“撇”, 如“帮”字下半部分“巾”中的第一个笔划“竖”, 根据隶书中这一笔的运笔特征, 应是一“撇”. 笔划“横折钩”变换成隶书后, 却成了“横折”. 如图10所示. 同样的规律, 在汉字“匝、两、内”等中也有反映. 因此, 特色笔划组“冂”, 在笔划替换时, 按笔划组做整体处理.



图10

3.8 部件内各笔划间拓朴关系的保持

两笔划间的拓朴关系有三种: 相离、相交、相切. 如图11所示, 其中相切是相交的特例.



图11 两笔划间的拓朴关系

拓朴特征是汉字的三大特征之一, 它是汉字识认和相互区别的依据. 在从黑体构架到隶书构架的笔划替换过程中, 由于隶书笔划与黑体笔划在斜率(对直线而言)或曲率(对曲线而言)上的差异, 有可能造成变换后隶书构架中隶书笔划间的位置关系与变换前原黑体构架中相应黑体笔划间的位置关系有微小出入, 致使改变了原有的拓朴关系, 如图12所示. 但由于拓朴关系是识别汉字的基础, 在一般情况下不应改变, 故必须对替换后改变了拓朴关系的笔

划进行调整,使之保持原有的拓朴关系.

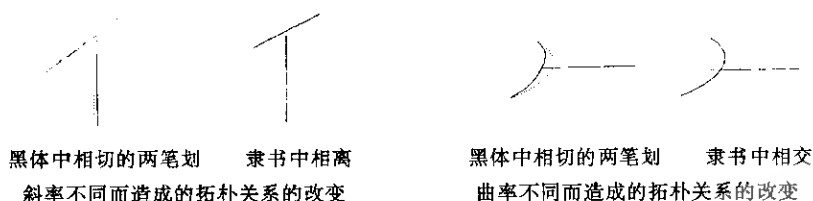


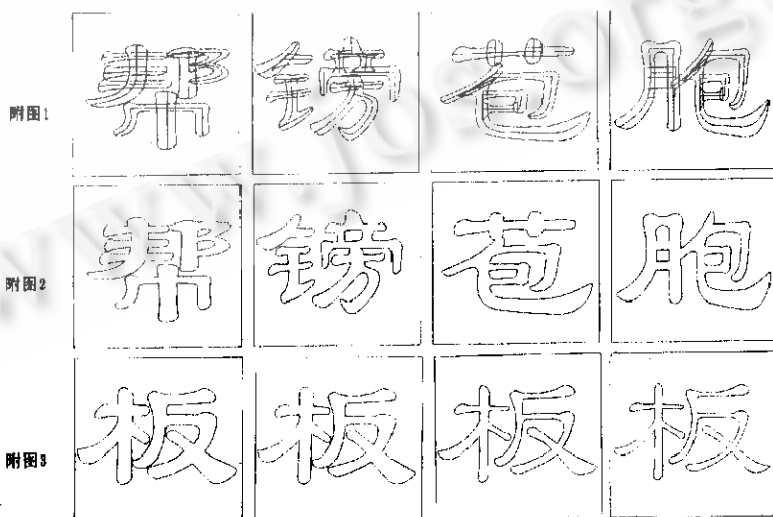
图 12

4 结 论

汉字字形从黑体到隶书的自动变体方法已在 286 微机上用 *Quick C 2.0* 实现. 整个变体过程由两部分程序实现: 第一部分程序实现的是黑体构架到隶书构架的自动变换, 程序量为 221K; 第二部分程序实现的是隶书笔划的柔性映射及轮廓字形的形成, 程序量为 171K. 另外还有用于交互式生成、修改基本笔划库和基本构架库以及其它的若干辅助程序.

变体的效率, 亦分两个步骤进行统计. 其中第一步从黑体构架变换到隶书构架, 平均每个字需 15.95 秒; 第二步从隶书构架到隶书轮廓字形的生成, 平均每个字需 67.15 秒. 这就是说, 从黑体按本方法变体到隶书的全过程, 平均每个字耗时 83.10 秒.

汉字已有丰富的字体, 它们陆续地都将被计算机化. 新的字体也还将会随着特定应用而诞生. 本文在异体字间的变体方面的尝试得到了成功. 附图 1 列出了使用本方法生成的基于笔划的隶书部分字样, 每一笔划的轮廓与中心线同时存在. 附图 2 则列出了经笔划求交处理后得到的完整的隶书轮廓的部分字样. 此外, 附图 3 给出了由本方法生成的不同粗细笔划的隶书字样对比. 笔划粗细通过对基本笔划先加工后得以控制. 这样一来, 本方法能生成的不再是一套隶书字体, 而是一组不同粗细的隶书字体. 它们分别适用于不同的应用场合, 比如, 根据字的不同大小尺寸, 选择相应粗细笔划的字体, 以保持美观.



参考文献

- 1 岳华等. 汉字黑体字形的一种衍生方法. 中文信息学报, 1995, 8(3).
- 2 武港山等. 等线体和圆头体曲线轮廓字形的自动生成系统. 中文信息学报, 1995, 8(4).
- 3 徐朝晖. 一种基于笔划映射的曲线轮廓汉字字形设计方法[硕士论文]. 南京大学, 1993.
- 4 任政. 隶书写法指南. 上海: 上海书店出版, 1991.
- 5 高晓. 一种从黑体到隶书的汉字字形自动变体方法[硕士论文]. 南京大学, 1994.

AN AUTOMATIC DERIVATION OF CHINESE CHARACTER TYPEFACE FROM HEI—TI TO LI—SHU

Gao Xiao Cai Shijie

(Department of Computer Science and Technology, Nanjing University, Nanjing 210093)

Abstract This paper introduces an automatic derivation of Chinese character typeface from Hei—Ti to Li—Shu, using the skeleton structure and basic strokes as the knowledge of a Chinese character. This method first implements the change from Hei—Ti skeleton structure to Li—Shu skeleton structure based on the stroke typeface characters. Then implements the flexible mapping of Li—Shu stroke and finally generates the outline of Li—Shu. This approach gives a way of automatic derivation of Chinese typefaces, and implements the automatic derivation from Hei—Ti to Li—Shu.

Key words Derivation, Hei—Ti, Li—Shu, skeleton, flexible mapping.