



第7章 汉字字形管理技术



内容

- ❖ 汉字字形概述
- ❖ 字形描述技术
- ❖ 字形的压缩和还原



7.1 汉字字形概述

- ❖ 汉字字形及其特点
- ❖ 字形、字型和字形库
- ❖ 字形描述技术及其种类



1.1 汉字字形及其特点

- ❖ 汉字字形指的是汉字形体结构的图像。
- ❖ 汉字字形的特点：
 - ✧ 汉字的字形呈方块形
 - ✧ 汉字从字形上可以分成各种结构
 - ✧ 汉字字形的分解方法和分解标准尚未统一，常用的字形分解方法一般从单字、字根、笔画（或笔形）和形素这四个层次入手。



ISO10646标准中对单字结构的定义

© ISO/IEC

ISO/IEC 10646-1 : 2000(E)

Table F.1: Properties of Ideographic description characters

Character Name: IDEOGRAPHIC DESCRIPTION CHARACTER ...	no. of DCs	IDS Acronym and Syntax	Relative positions of DCs	Example of IDS	IDS example represents:
LEFT TO RIGHT	2	IDC-LTR D ₁ D ₂		母	母
ABOVE TO BELOW	2	IDC-ATB D ₁ D ₂		天	天
LEFT TO MIDDLE AND RIGHT	3	IDC-LMR D ₁ D ₂ D ₃		言	言
ABOVE TO MIDDLE AND BELOW	3	IDC-AMB D ₁ D ₂ D ₃		从	从
FULL SURROUND	2	IDC-FSD D ₁ D ₂		巷	巷
SURROUND FROM ABOVE	2	IDC-SAV D ₁ D ₂		門	門
SURROUND FROM BELOW	2	IDC-SBL D ₁ D ₂		山	山
SURROUND FROM LEFT	2	IDC-SLT D ₁ D ₂		匚	匚
SURROUND FROM UPPER LEFT	2	IDC-SUL D ₁ D ₂		舞	舞
SURROUND FROM UPPER RIGHT	2	IDC-SUR D ₁ D ₂		去	去
SURROUND FROM LOWER LEFT	2	IDC-SLL D ₁ D ₂		交	交
OVERLAID	2	IDC-OVL D ₁ D ₂		从	从

* NOTE - D₁ and D₂ overlap each other. This diagram does not imply that D₁ is on the top left corner and D₂ is on the bottom right corner.



1.2 字形、字型和字形库

- ❖ 字形（glyph）：字符的一个实例。
- ❖ 有的时候，几个字符可以合成一个字形，例如美元符号：
 - ✧ 一个“S”形状加上一根竖线：\$
 - ✧ 一个“S”形状加上一根割断的竖线：\$
 - ✧ 一个“S”形状加上两根竖线：\$
 - ✧ 一个“S”形状加上两根割断的竖线：\$



1.2 字形、字型和字形库

- ❖ 汉字的字形更加复杂，我国制定了一系列的国家标准来严格定义各种字符集中字符的字形。
 - ✧ 《信息交换用汉字32 × 32点阵字模集及数据集》：规定了GB2312—80字符集中各个字符的各种字形。
 - ✧ 《信息技术——通用多八位编码字符集48点阵字形》：覆盖了GBK字符集的所有字符并规定了这些字符的点阵字形。



1.2 字形、字型和字形库

❖ ISO和Unicode联盟对字符、字形和字形图像的定义。

Terminology	ISO	Unicode ^a
Character	A member of a set of elements used for the organization, control, or representation of data. ^b An atom of information with an individual meaning, defined by a character repertoire. ^c	(1) The smallest component of written language that has semantic value; refers to the meaning and/or shape, rather than a specific shape, (see also <i>glyph</i>) though in code tables some form of visual representation is essential for the reader's understanding.
Glyph	A recognizable abstract graphical symbol which is independent of any specific design. ^c	(1) An abstract form that represents one or more glyph images. (2) A synonym for <i>glyph image</i> .
Glyph image	An image of a glyph, as obtained from a glyph representation displayed on a presentation surface. ^c	The actual, concrete image of a glyph representation having been rasterized or otherwise imaged onto some display surface.



1.2 字形、字型 and 字形库

- ❖ 字体 (typeface) : 一个字形或一个字符集印刷出来的样式。
- ❖ 字型 (font) : 某一个字体的实例, 如某个特定的磅值。
- ❖ 西文的印刷字型有:
 - ☞ serif
 - ☞ sans serif
 - ☞ script
- ❖ 中文使用的主要字型有:
 - ☞ Song (宋体)
 - ☞ Fangsong (仿宋体)
 - ☞ Kai (楷体)
 - ☞ Hei (黑体)



1.2 字形、字型 and 字形库

西文字型	中文字型
Serif	Song (宋体)
Sans serif	Hei (黑体)
Script	Kai (楷体)
Other	Fangsong (仿宋体)

本地化过程中，西文字型和中文字型的对应关系



1.2 字形、字型和字形库

- ❖ 字形库中存储的是某个集合所有字符的字形信息。
- ❖ 实际上，字形库是一个计算机的存储器系统，它的设计、管理和使用原则上与计算机存储器系统是一致的。



1.3 字形描述技术及其种类

- ❖ 字形描述技术是以**数字代码描述字的形状**，在输出版面的时候，由计算机将代码还原，恢复字原来的形状，由输出设备输出。字形描述技术也被称为**字形压缩还原技术**。
- ❖ 字形描述技术的水平高低，不仅决定了输出字的速度、字形美观和输出效果，同时也影响到：
 - ❧ 字库存储空间
 - ❧ 字形还原的计算量



1.3 字形描述技术及其种类

❖ 字形描述技术描述一般可分为：

- ❧ 点阵字形描述技术
- ❧ 轮廓矢量字形描述技术
- ❧ 曲线轮廓字形描述技术



2. 字形描述技术

- ❖ 点阵字形描述技术
- ❖ 轮廓矢量字形描述技术
- ❖ 曲线轮廓字形描述技术
- ❖ 其他字形描述技术



2.1 点阵字形描述技术

- ❖ 汉字点阵的规模一般是： $m \cdot 2^n \times m \cdot 2^n$
- ❖ 在一般情况下， 16×16 点阵和 24×24 点阵使用的比较多。随着激光印字机、激光照排机、图文记录机等输出设备分辨力的提高，又有了 48×48 、 96×96 、 128×128 、 256×256 、 512×512 、 1024×1024 的点阵字库。



2.1 点阵字形描述技术

❖ 16 × 16点阵字形描述方法介绍

- ✧ 把一个方块横向分成16格，纵向也分成16格，形成256个小方格，即该矩阵有256个“点”。每一个点可以有黑、白两种颜色。
- ✧ 用这样的点阵覆盖到汉字上，凡是与笔画重叠的点，规定它是黑色；凡是不与笔画重叠的点，规定它是白色。

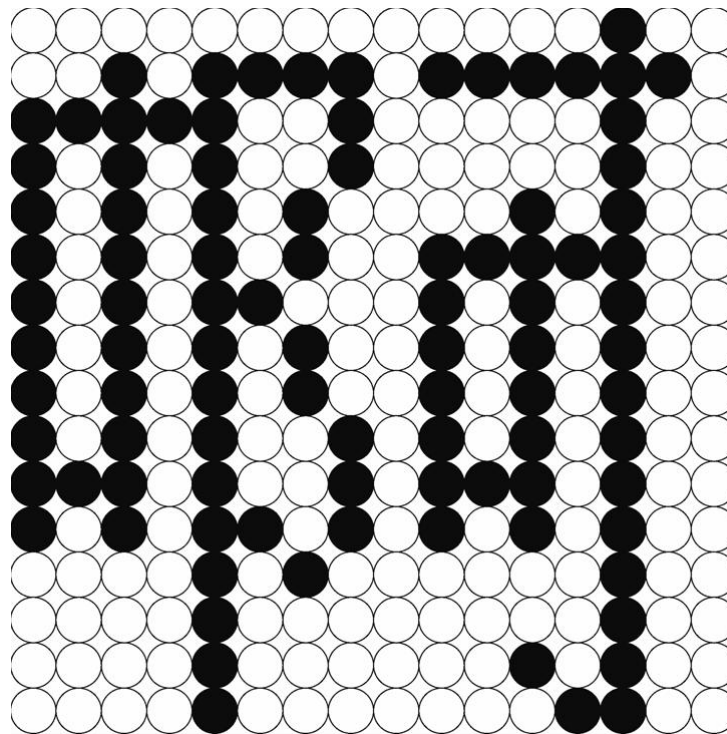


2.1 点阵字形描述技术

- ❖ 用二进制数的“1”和“0”来分别表示图中黑色和白色的点。一个 16×16 的点阵字形就可以用一串二进制数（256位二进制数）来表示。这样的方法叫做点阵的数字化。



2.1 点阵字形描述技术



汉字“啊”的点阵字形

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	0
1	1	1	1	1	0	0	1	0	0	0	0	0	0	1	0	0	0
1	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0
1	0	1	0	1	0	1	0	0	0	0	1	0	1	0	1	0	0
1	0	1	0	1	0	1	0	0	1	1	1	1	1	1	0	0	0
1	0	1	0	1	1	0	0	0	1	0	1	0	1	0	1	0	0
1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0	0
1	0	1	0	1	0	1	0	0	1	0	1	0	1	0	1	0	0
1	0	1	0	1	0	0	1	0	1	0	1	1	1	0	1	0	0
1	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	0
0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0
0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0

汉字“啊”点阵字形数字化表示



2.1 点阵字形描述技术

- ❖ 一个汉字点阵字形的存储量：
 - ✧ 16×16 : 256位，即32个字节
 - ✧ 24×24 : 72个字节
 - ✧ 32×32 : 128个字节



2.1 点阵字形描述技术

- ❖ 横向点阵：每个点阵数字代码均描述汉字横向点的情况。数字代码是顺序从左到右、自上而下扫描点阵各点状态而形成的。一般适用于显示器。
- ❖ 纵向点阵：每个点阵数字代码均描述汉字纵向点的情况。数字代码是顺序从上到下、自左至右扫描点阵各点状态而形成的。一般适用于打印机。



2.1 点阵字形描述技术

第0字节	第1字节
第2字节	第3字节
第4字节	第5字节
第6字节	第7字节
第8字节	第9字节
第10字节	第11字节
第12字节	第13字节
第14字节	第15字节
第16字节	第17字节
第18字节	第19字节
第20字节	第21字节
第22字节	第23字节
第24字节	第25字节
第26字节	第27字节
第28字节	第29字节
第30字节	第31字节

16 × 16横向点阵的情况



2.1 点阵字形描述技术

❖ 汉字“啊”用横向点阵来表示，其数字代码是：

00H、04H、2FH、7EH、F9H、04H、
A9H、04H、AAH、14H、AAH、7CH、
ACH、54H、AAH、54H、AAH、54H、
A9H、54H、E9H、74H、ADH、54H、
0AH、04H、08H、04H、08H、14H、
08H、0CH



2.1 点阵字形描述技术

第1字节	第0字节
第3字节	第2字节
第5字节	第4字节
第7字节	第6字节
第9字节	第8字节
第11字节	第10字节
第13字节	第12字节
第15字节	第14字节
第17字节	第16字节
第19字节	第18字节
第21字节	第20字节
第23字节	第22字节
第25字节	第24字节
第27字节	第26字节
第29字节	第28字节
第31字节	第30字节

16 × 16纵向点阵的情况



2.1 点阵字形描述技术

❖ 汉字“啊”用纵向点阵来表示，其数字代码是：

3FH、F0H、20H、20H、7FH、F0H、20H、
00H、7FH、FFH、42H、10H、4DH、88H、
70H、70H、00H、00H、47H、F0H、44H、
20H、4FH、F2H、44H、01H、FFH、FFH、
40H、00H、00H、00H

❖ 横向点阵和纵向点阵可以互相转换。



2.1 点阵字形描述技术

❖ 各种精度的点阵字形其特点和用途：

✧ 16 × 16（简易型）

- ❖ 通常采用单线体。
- ❖ 不太美观。
- ❖ 需要进行简化处理。
- ❖ 只能适用于低分辨率的显示终端用。



2.1 点阵字形描述技术

❖ 各种精度的点阵字形其特点和用途：

🌀 24 × 24（普通型）

- ❖ 大多数汉字都可以用双线体。
- ❖ 可以做到笔画齐全，横细直粗，具有宋体字的风格，字形比较准确美观。
- ❖ 适用于普通打印机和较好的显示输出之用。



2.1 点阵字形描述技术

❖ 各种精度的点阵字形其特点和用途：

🌀 32 × 32（提高型）

- ❖ 采用多线体
- ❖ 比24 × 24点阵字形更好的体现宋体字的风格。
- ❖ 可以完整的表达笔画，可表现出一定程度的笔锋变化，撇捺自然婉转，舒畅流利。
- ❖ 适用于较高质量的打印文件使用。



2.1 点阵字形描述技术

❖ 各种精度的点阵字形其特点和用途：

✎ 其他更高精度的点阵字形

- ❖ 48×48 点阵字形，将可以更好的描绘宋体、仿宋体、楷书、黑体等多种风格的汉字，也可以较满意的放大字形。但由于点阵大，占用的存储空间大，目前固化有困难，故只能存放在磁盘中。这样，汉字在显示时的速度将受到一定的影响，尤其是在高速激光打印时，速度的矛盾更加尖锐，系统的成本也比较高。
- ❖ 64×64 以上的点阵字形，一般来说，要采用压缩还原技术，成本很高，较多的用在照排系统中。

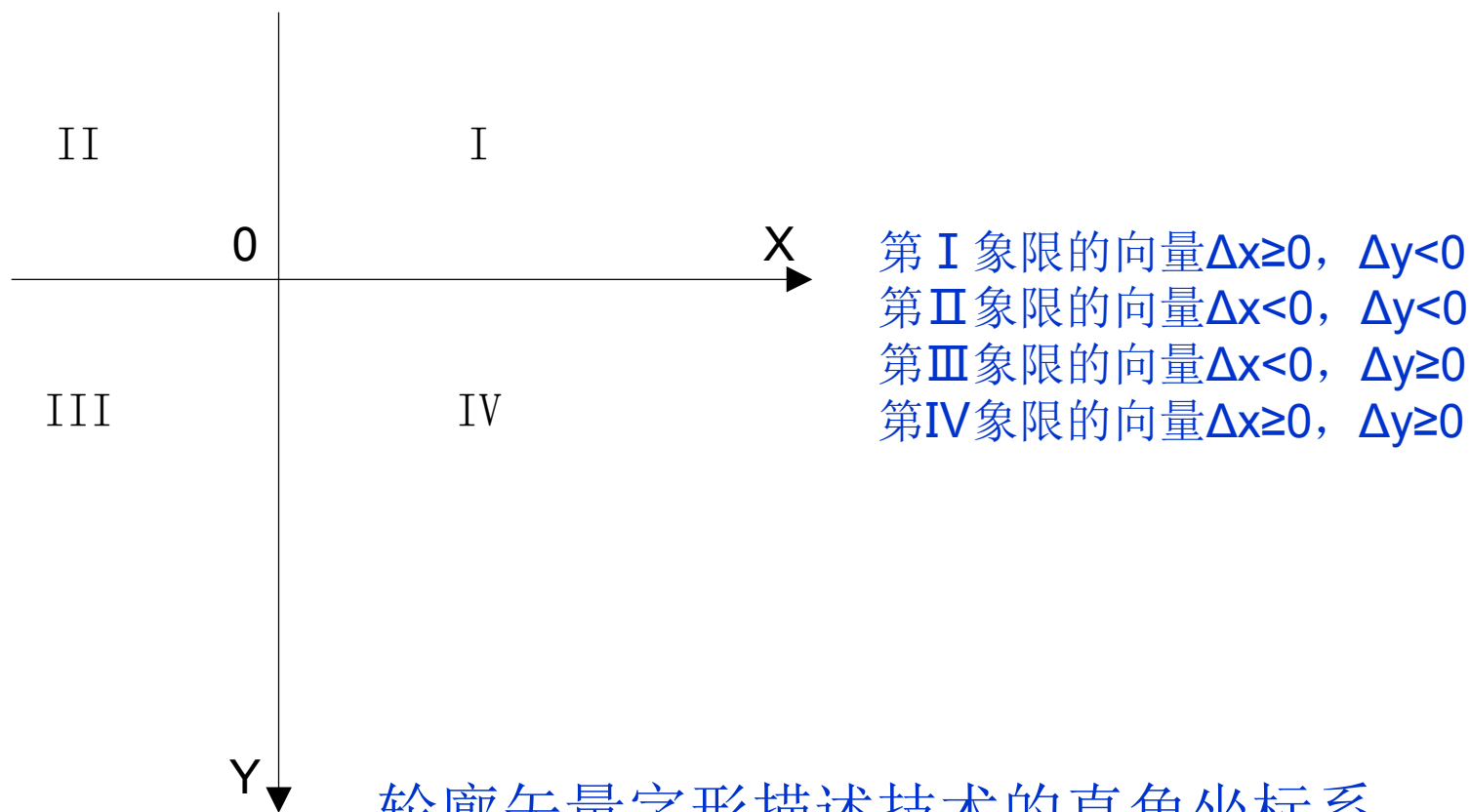


2.2 轮廓矢量字形描述技术

- ❖ 轮廓矢量字库的出现是为了改进点阵字库锯齿状轮廓，同时也是因为绘图机用笔输出的需要而发展起来的一种字形描述技术。
- ❖ 轮廓矢量的字形描述技术的核心是用多条直线逼近字形，并进行封闭处理，字形效果明显改观。



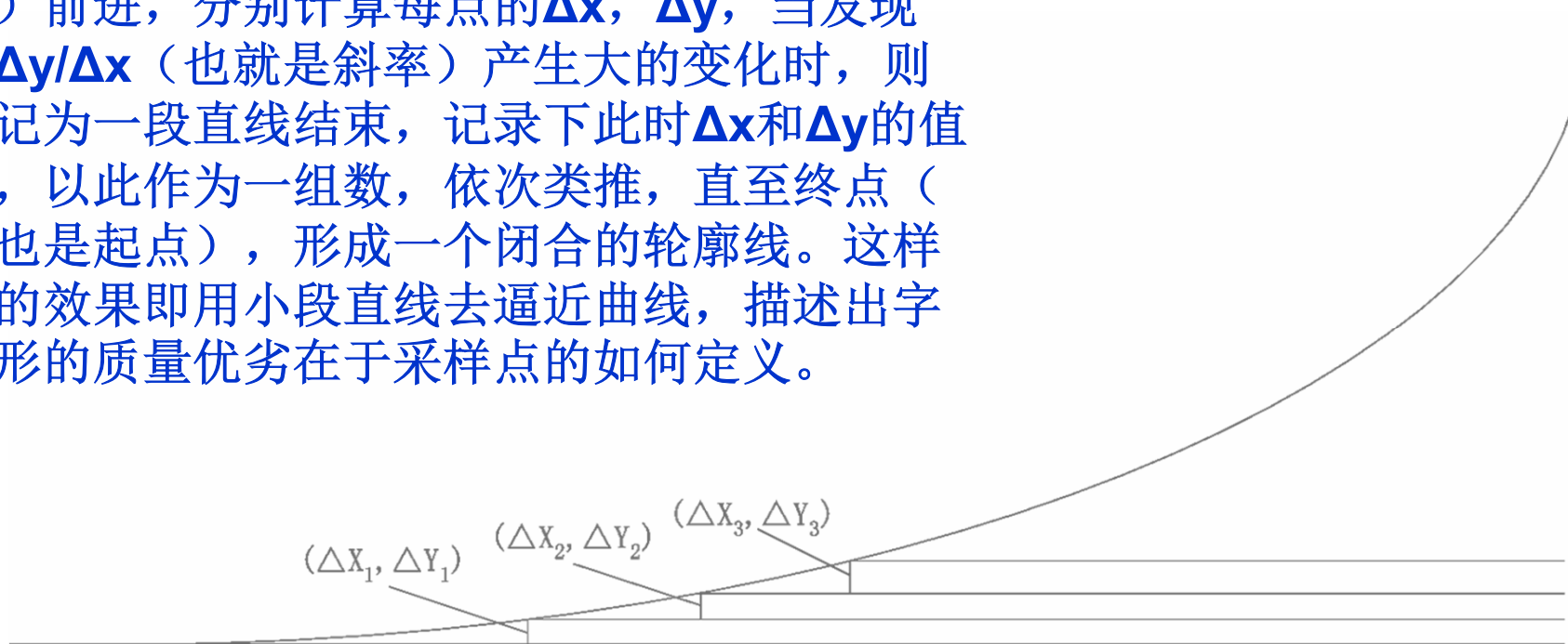
2.2 轮廓矢量字形描述技术





2.2 轮廓矢量字形描述技术

给出起始向量的起点坐标 (x, y) ，然后分段沿轮廓线按照一定方向（顺时针或逆时针）前进，分别计算每点的 Δx , Δy ，当发现 $\Delta y/\Delta x$ （也就是斜率）产生大的变化时，则记为一段直线结束，记录下此时 Δx 和 Δy 的值，以此作为一组数，依次类推，直至终点（也是起点），形成一个闭合的轮廓线。这样的效果即用小段直线去逼近曲线，描述出字形的质量优劣在于采样点的如何定义。





2.2 轮廓矢量字形描述技术

- ❖ 轮廓字库解决了点阵字库缩放后失真的问题。这种字形描述技术对汉字横平竖直的字形有很好的描述效果，但对于撇、捺、点，以及圆弧装饰效果的字体，仍然不能十分逼近字形，而且字体放大之后，轮廓仍然具有折线感。
- ❖ 在大的点阵做成的矢量轮廓字库缩小到很小或小的点阵做成的矢量字放大得很大时，都会带来失真。
- ❖ 因此大号字用 256×256 点阵为基础的矢量轮廓，而小号字仍用 96×96 或 128×128 点阵为基础的矢量轮廓。



2.3 曲线轮廓字形描述技术

- ❖ 曲线轮廓字库是以二次曲线和三次曲线逼近字形轮廓的字形描述方法。其中以 **PostScript** 字形描述技术为代表。
- ❖ **PostScript** 字形描述技术是用美国 **Adobe** 公司的 **PostScript** 页面描述语言来描述字形的一种技术。其特点是以贝塞尔曲线（**Bezier curve**）拟合文字形状，比较完整地保留了文字原有的字形信息。



2.3 曲线轮廓字形描述技术

❖ 解决字形质量的失真问题:

- ❧ 特征参数: 对用直线、二次曲线或三次曲线描述字形轮廓的字形描述技术的一种补充。
- ❧ 智能复原: 依靠字形复原软件的判断和处理来保证字形复原的质量, 而字形描述本身不包含很多的特征信息。



2.3 曲线轮廓字形描述技术

❖ PostScript Type 1语言中的特征参数:

☞ 字符级特征参数:

- ❖ hstem/vstem: 给出水平/垂直笔段的位置和宽度。
- ❖ hstem 3/vstem 3: 给出等距等宽的三个平行笔段的位置和宽度。

☞ 字库级特征参数:

- ❖ StdHW, StdVW, StemSnapH, StemSnapV: 给出字库中字符的标准笔画宽带, 或者笔画宽度的跳跃变化, 参数为宽度值的数组。
- ❖ BlueValues: 给出西文字形的准线位置, 参数为准线位置的数组。



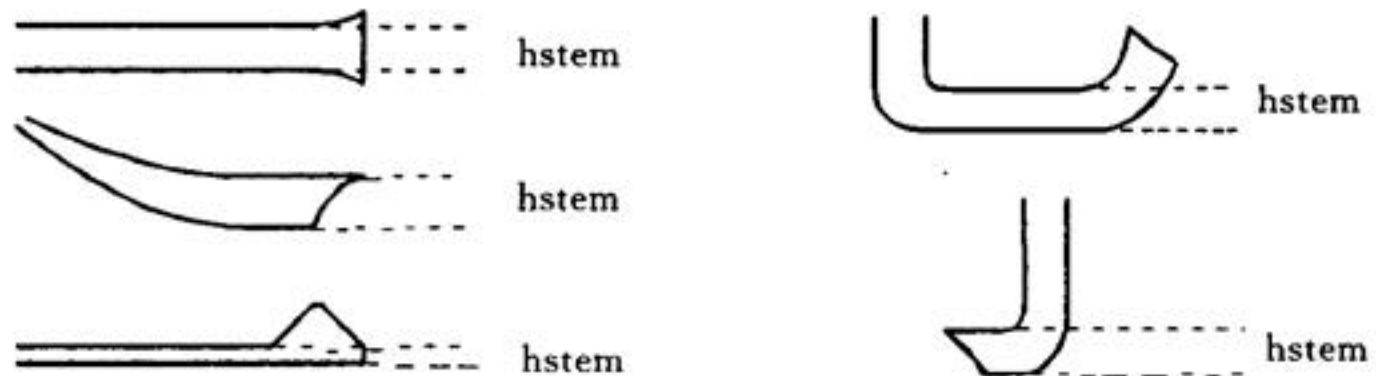
2.3 曲线轮廓字形描述技术

❖ 给汉字添加特征参数

- ✧ 将汉字分解为一些简单的笔画，再针对不同笔画的几何特征给出相应的 **hstem** / **vstem**。
- ✧ 对 **hstem** / **vstem** 所需参数加以测量，**hstem** 的参数是 y 、 Δy ，**vstem** 的参数是 x 、 Δx ，它们分别表示水平/垂直笔段在字符空间上的位置和宽度。



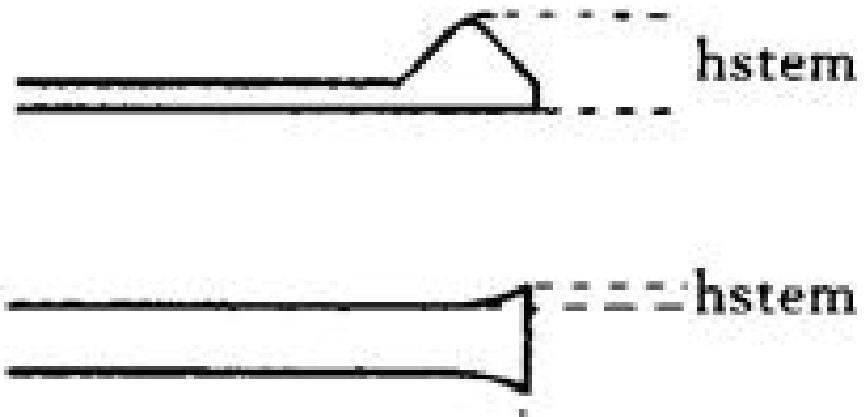
2.3 曲线轮廓字形描述技术



部分笔画水平段的情况



2.3 曲线轮廓字形描述技术



笔锋高度的测量



2.3 曲线轮廓字形描述技术

❖ 字库中每个汉字的特征参数生成完毕以后，还要添加合适的字库级特征参数，主要包括：

✧ 笔画标准宽度StdHW/StdVW

✧ 笔画宽度的跳跃变化
StemSnapH/StemSnapV



2.3 曲线轮廓字形描述技术

- ❖ 比较有影响的PostScript西文字库有：
 - ✧ Type 1: Adobe公司自行设计的字体
 - ✧ Type 3: 其他厂商设计的字体
 - ✧ Type 5: 打印机ROM中携带的字体
- ❖ 国内用PostScript语言描述的轮廓字体只有以Type 1 和Type 3的格式做出的字库。



2.3 曲线轮廓字形描述技术

- ❖ 用PostScript技术描述汉字时，在字符笔画过大、字又太细长时，软件为了要显示所有黑色笔画就会忽略笔画间的空白，造成笔画重叠、比例变形等问题。
- ❖ 对于低分辨力的输出设备（如显示器和针式打印机），这种字库的输出效果反倒不如点阵字库优越，即使增加大量的特征参数描述也得不到较大的改善。因此也采用字库中包含点阵字库用于输出小字号文字的变通方法。



2.4 其他字形描述技术

❖ TrueType字形技术

- ✧ 微软公司协助苹果公司开发并于1990年底推出TrueType字库。这种字库放弃以三次曲线逼近字形轮廓的优点，而采用二次曲线逼近字形轮廓。TrueType字库有精心设计的特征参数处理。

❖ Folio字形技术

- ✧ 在Unix作站的领域中，有Sun公司的Folio字库技术。由于Sun工作站在美国广泛用于社团和专业用的出版系统，因此Folio F3字形描述格式也是值得一提的字形描述技术。

❖ 智能字形技术（Intellifont）

- ✧ 由爱克发计算机国际公司开发，用于惠普系列激光印字机。



3. 字形的压缩和还原

- ❖ 什么是字形的压缩和还原
- ❖ 字形点阵的压缩和还原
- ❖ 轮廓矢量字形的压缩方法
- ❖ 其他字形压缩技术



3.1 什么是字形的压缩和还原

- ❖ 对于正式出版物来说，文字分辨率需29.2线/毫米以上。因此，最常用的五号汉字点阵为 108×108 ，特大号字则为 576×576 。
- ❖ 不同精度的点阵表示一个汉字需要的字节数：

48×48	288字节
96×96	1152字节
128×128	2048字节 (2KB)
256×256	8192字节 (8KB)
512×512	32768字节 (32KB)
1024×1024	131072字节 (128KB)



3.1 什么是字形的压缩和还原

- ❖ 对于 512×512 点阵，存放GB2312-80字符集的所有汉字，四种字体（宋体、楷体、黑体、仿宋体），字库容量需要：

$$\approx (512/8) \times 512 \times 6763 \times 4 = 845.375\text{MB}$$

- ❖ 对于 1024×1024 点阵，存放GB2312-80字符集的所有汉字，四种字体（宋体、楷体、黑体、仿宋体），字库容量需要

$$\approx (1024/8) \times 1024 \times 6763 \times 4 = 3381.5\text{MB}$$



3.1 什么是字形的压缩和还原

- ❖ 字形信息压缩技术应满足以下几方面要求：
 - ❧ 字形信息压缩技术不允许信息压缩后丢失或发生错误，也就是说压缩信息能正确再现原字形信息，否则在文字复原时将会引起字形畸变。
 - ❧ 字形信息压缩过程应当简便易行，并且可用专门的软件和硬件自动完成，便于用户独立进行缺字补充的操作。
 - ❧ 由文字字形压缩信息向最终输出的点阵字形信息的转换应当迅速可靠。
 - ❧ 压缩效果显著。



3.1 什么是字形的压缩和还原

- ❖ 追求**压缩效果**和**文字复原速度**这两个指标是相互冲突的，因为**压缩倍率越高**，字形**复原过程就越复杂**，复原工作量也越大，文字还原速度也就越慢。
- ❖ 事实上，任何字形信息的压缩方法都是以损失文字复原速度和输出速度为代价而换来存储空间的节约的。
- ❖ 评价一种字形压缩技术的优劣，不仅要看信息压缩倍率的高低，而且还要看是否能快速复原。



3.2 字形点阵的压缩和还原

❖ 黑白段表示法

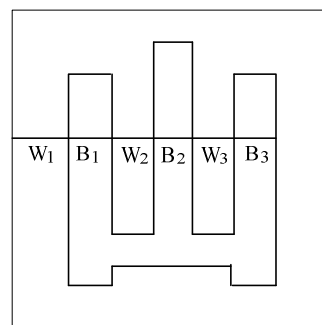
- 黑白段表示法是考虑到汉字笔画中黑白连续变化的特点而产生的压缩方法。
- 压缩时计算机逐行地数取点阵中的空白点数（白段）和非空白点数（黑段），按一定格式相间地记录下一串数字构成行点阵信息。
- 同样也可用重复行数来压缩相邻的重复行信息。
- [行标志][重复行数][白段][黑段]...[白段][黑段]

* N W₁ B₁ ... W_n B_n



3.2 字形点阵的压缩和还原

❖ 黑白段表示法



❖ 实例：英国蒙纳汉字激光照排系统。



3.2 字形点阵的压缩和还原

❖ 线性增量表示法

- ❧ 采用黑、白段信息格式，保留上述的竖直压缩，增加一种新的压缩格式。
- ❧ 线性增量表示法用以处理汉字笔画中大量存在的斜线。
- ❧ 在一行黑、白段记录信息的后面再注明线段的增量，这样下一行的黑、白段长度在上一行的基础上按增量的大小作相应的变化。



3.2 字形点阵的压缩和还原

❖ 线性增量表示法

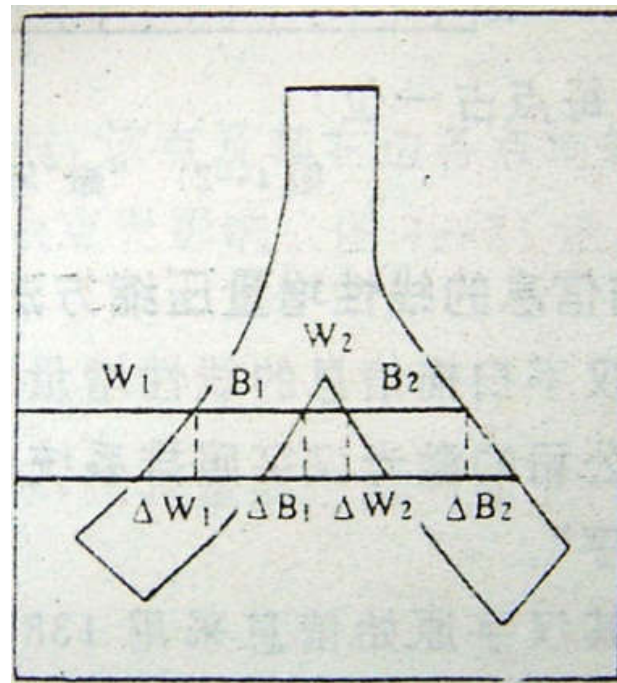
[行标志][重复行数][白段][白段增量][黑段][黑段增量]
** N W₁ ΔW₁ B₁ ΔB₁
...[白段][白段增量][黑段][黑段增量]
... W_n ΔW_n B_n ΔB_n

- ❧ 用增量记录的行数要超过两行（n大于2），压缩才有意义。
- ❧ 第1行是不考虑增量值的黑、白段信息，最后一行是加上相应增量后的黑、白段信息，中间的n-2行的黑白段长段均在上一行的长度基础上，加上线性增量来补齐的。



3.2 字形点阵的压缩和还原

❖ 线性增量表示法





3.2 字形点阵的压缩和还原

- ❖ 思考：考虑 1024×1024 点阵字库，如采用黑白段表示法得到一行信息：“* 2 1024 0”，它表示什么？
- ❖ 如果相邻行不相同的概率为50%，平均每行有5个黑段，则一个汉字的字形信息有：
 $(1+1+10*2)*1024*(100\%-50\%)=11K$
- ❖ 国内早期在蒙纳激光照排系统中使用一种 1380×1380 点阵的字库，1种字体就要约2000MB的存储量。
- ❖ 采用这两种压缩方法以后，这个系统的黑体、宋体、仿宋体、楷体4种字体的字库总存储量约为65MB。



3.3 轮廓矢量字形的压缩方法

- ❖ 规则笔画
 - ☞ 横、竖、折。
 - ☞ 规则笔画由直线段、起笔笔锋、收笔笔锋和转折笔锋等构成。
- ❖ 直线段可以有轻微的倾斜。
- ❖ 笔锋总共可以分成7种，因此在宋体横这一笔画的压缩表示中，用3位表示收笔笔锋的号码（0—6），其余的信息将指示横的起始点的x、y坐标，横的长度、宽度以及有无起笔笔锋等。这样宋体的一笔横需3—4字节便可精确描述而丝毫不失真。
- ❖ 笔画竖也是如此，起笔笔锋用3位信息来表示，收起笔锋用2位信息来表示，其余的信息则表示起点坐标、竖的长度和宽度，此外还用1个字节来详细描述笔画左、右倾斜的情况，这样描述宋体的竖需要5个字节。



3.3 轮廓矢量字形的压缩方法

❖ 不规则笔画

❧ 点、撇、捺、钩...

❧ 对不规则笔画，用向量折线对轮廓曲线作一般性的描述。

❧ 为了减少数位的浪费，使压缩信息更加紧凑，采用增设控制字节的方法，按下述规则对之进行装配：



3.3 轮廓矢量字形的压缩方法

- ❖ 若控制字节的头两位为11，则该字节其余6位与下一节合起来，指示不规则笔画起始点的x、y坐标。

7	6	5	4	3	2	1	0
1	1	笔画起始点的x坐标					
x_7	笔画起始点的y坐标						



3.3 轮廓矢量字形的压缩方法

- ❖ 若控制字节的头两位为10，表示所指示的向量的 Δx 、 Δy 的绝对值中至少有一个大于等于16，此时控制字节的末6位与下一字节合起来表示该向量的 Δx 、 Δy 的数值，数值以原码表示。

7	6	5	4	3	2	1	0
1	0	向量的 Δx					
x_7	向量的 Δy						



3.3 轮廓矢量字形的压缩方法

- ❖ 若控制字节的头两位为01，表示后面N个向量的 Δx 、 Δy 绝对值均小于8，N由控制字节的最末5位指示。后面N个向量中的每个向量用一个字节表示，其中2位指示该向量所处的象限，3位指示 Δx 的绝对值，3位指示 Δy 的绝对值。

7	6	5	4	3	2	1	0
0	1	向量的个数N					
象限		Δx			Δy		
N		⋮			⋮		
象限		Δx			Δy		



3.3 轮廓矢量字形的压缩方法

- ❖ 若控制字节的头两位为00，表示后面N个向量处于同一象限，且 Δx 、 Δy 的绝对值均小于16。N由控制字节的最末4位表示，控制字节的中间2位指明象限，后面N个向量都由一个字节表示，其中4位指示 Δx 的绝对值，4位指示 Δy 的绝对值。

7	6	5	4	3	2	1	0
0	0	象限		向量的个数N			
N		$ \Delta x $		$ \Delta y $			
		\vdots		\vdots			
		$ \Delta x $		$ \Delta y $			



3.3 轮廓矢量字形的压缩方法

- ❖ 在采用了上述的进一步压缩的措施之后，1个五号汉字字形信息平均需要**120**字节来记录，而记录1个五号汉字字形点阵信息大约需用**1400**多个字节，因此单字的平均压缩倍率大约为**12**倍。
- ❖ 由于五号字最常用，且非变倍字形的恢复速度也最快，故系统字库中只存入五号字的字形压缩信息，而其余字号均由计算机通过五号字字模变倍得到。
- ❖ 目前的计算机汉字出版系统上中宋体、黑体、仿宋体、楷体、报宋体、小标题宋体几种字模的压缩信息，存储量约为**5MB**，字模总体压缩倍率在**500**倍左右。



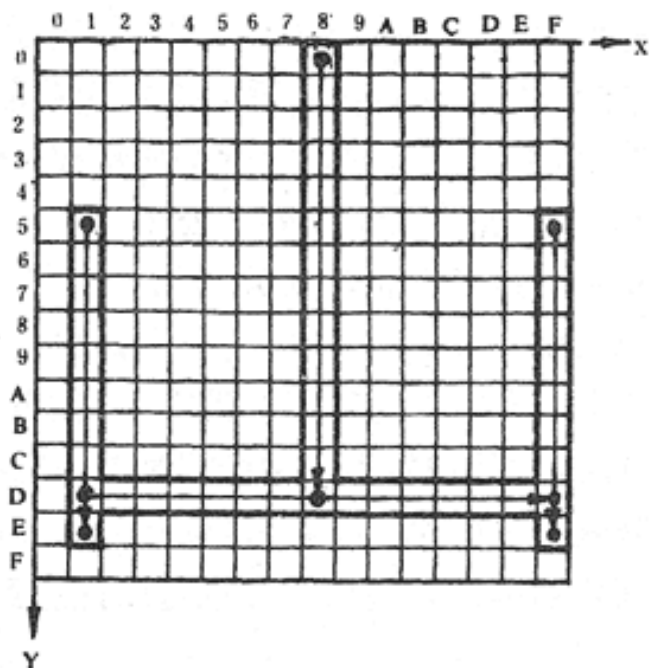
3.3 轮廓矢量字形的压缩方法

- ❖ 字形压缩信息与最终输出的点阵字形信息，这两种数据之间存在着“结构冲突”。
 - ❧ “次序冲突”：字模压缩信息以笔画为序，规则笔画在前，不规则笔画在后，一笔信息完毕再有另一笔信息；而输出点阵信息则是从左至右，一行完了再开始新的一行，二者的次序关系完全不同。
 - ❧ “边界冲突”：字模压缩信息以笔画划分内部信息的边界，而最终要输出的点阵信息以行划分边界，二者的边界完全不同。



3.4 其他字形压缩技术

- ❖ 骨架向量法：又称笔画骨架法。用能够表示汉字笔画特征的线段，作为组成汉字字形表示法。



这个“山”字的字形用8个字节即可表达

可用下列数据来表示：

(15, 1E)、(80, 8D)

(F5, FE)、(1D, FD)



3.4 其他字形压缩技术

- ❖ 部件组字法：抽取汉字的共性，以最少量的部件和笔画作为基本字元来组织出尽可能多的汉字字形，这就是所谓的组字技术。组字技术并非信息压缩技术。
- ❖ 汉字字形压缩法还有：无失真向量拟合法、曲线参数拟合法、笔画函数叠加法、轮廓特征分析法、复合压缩法、递推偏差拟合法、折线法、叠加法字形压缩、哈夫曼综合处理法等。



作业

❖ P.160 1-7