



中华人民共和国国家标准

GB/T 21049—2007

汉 信 码

Chinese-sensible code



2007-08-23 发布

2008-02-01 实施

中华人民共和国国家质量监督检验检疫总局
中国国家标准化管理委员会 发布



中 华 人 民 共 和 国
国 家 标 准
汉 信 码

GB/T 21049—2007

*

中国标准出版社出版发行
北京复兴门外三里河北街16号

邮政编码:100045

网址 www.spc.net.cn

电话:68523946 68517548

中国标准出版社秦皇岛印刷厂印刷

各地新华书店经销

*

开本 880×1230 1/16 印张 4.5 字数 130 千字
2007年10月第一版 2007年10月第一次印刷

*

书号:155066·1-30005 定价 44.00 元

如有印装差错 由本社发行中心调换

版权专有 侵权必究

举报电话:(010)68533533



GB/T 21049-2007

前 言

本标准规定了一种矩阵式二维条码——汉信码的码制以及编译码方法。本标准中对汉信码的符号结构、信息编译码方法、纠错编译码算法、信息排布方法、参考译码算法等内容进行了详细的描述,汉信码可高效表示 GB 18030—2000《信息技术 信息交换用汉字编码字符集 基本集的扩充》中的汉字信息,并具有数据容量大、抗畸变和抗污损能力强、外观美观等特点,适合于在我国各行业广泛应用。

本标准的附录 A、附录 B、附录 C、附录 D、附录 E、附录 F、附录 H 是规范性附录,附录 G 是资料性附录。

本标准由中国标准化研究院提出。

本标准由中国标准化研究院中国物品编码中心归口。

本标准起草单位:中国物品编码中心、北京网路畅想科技发展有限公司、北京意锐新创科技有限公司。

本标准主要起草人:张成海、赵楠、黄燕滨、罗秋科、王毅、张铎、王越。

引 言

本标准的发布机构提请注意如下事实,声明符合本标准时,可以使用涉及 5.2 节、第 6 章、第 9 章、第 10 章、附录 E 中有关的相关专利。

本标准的发布机构对于专利的范围、有效性和验证资料不提出任何看法。

专利持有人已向本标准的发布机构保证,任何申请人皆可免费获得使用授权许可。在这方面,该专利持有人的声明已在本标准的发布机构备案。有关资料可从以下地址获得:

专利所有人	中国物品编码中心
地址	北京市北三环中路 3 号双全大厦 4 层
邮政编码	100029
网址	http://www.ancc.org.cn
联系人	王毅
联系电话	010-62358298
传真	010-82029374
E-mail	wangy@ancc.org.cn

请注意除上述已经识别出的专利外,本标准的某些内容有可能涉及专利。本标准的发布机构不应承担识别这些专利的责任。

目 次

前言 I

引言 II

1 范围 1

2 规范性引用文件 1

3 术语与定义 1

4 约定 3

5 符号特性与符号结构 3

6 数据编码与符号表示 11

7 符号印制 23

8 符号质量 23

9 译码过程 23

10 汉信码的参考译码算法 24

附录 A (规范性附录) 汉信码数据容量 30

附录 B (规范性附录) 汉信码信息容量 36

附录 C (规范性附录) 纠错译码步骤 47

附录 D (规范性附录) 纠错码生成多项式 48

附录 E (规范性附录) 汉信码符号各版本的纠错特性 50

附录 F (规范性附录) 功能信息 61

附录 G (资料性附录) 汉信码符号印制与扫描的用户导则 63

附录 H (规范性附录) 汉信码印制质量——码制特殊指标 64

汉 信 码

1 范围

本标准规定了汉信码符号的符号结构、信息编译码方法、纠错编译码算法、信息排布方法、参考译码算法、符号质量要求等技术内容。

本标准适用于自动识别和数据采集。

2 规范性引用文件

下列文件中的条款通过本标准的引用而成为本标准的条款。凡是注日期的引用文件,其随后所有的修改单(不包括勘误的内容)或修订版均不适用于本标准,然而,鼓励根据本标准达成协议的各方研究是否可使用这些文件的最新版本。凡是不注日期的引用文件,其最新版本适用于本标准。

GB/T 11383 信息处理 信息交换用八位代码结构和编制规则(GB/T 11383—1989, idt ISO 4873:1986)

GB/T 12905 条码术语

GB 18030 信息技术 信息交换用汉字编码字符集 基本集的扩充

ISO/IEC 15415 信息技术 条码符号质量的检验 二维条码符号

AIM 国际技术规范 扩展解释:第一部分:识别方案与协议(简称“AIM ECI 规范”)

3 术语与定义

GB/T 12905 中规定的术语和下列术语与定义适用于本标准。

3.1

功能图形 **function pattern**

符号中用于符号定位与特征识别的特定图形。

3.2

位置探测图形 **position detection pattern**

组成汉信码寻像图形的4个特定的图形。

3.3

位置探测中心

汉信码位置探测图形 3×3 深色块的中心。

3.4

校正图形 **alignment pattern**

位于汉信码符号内,用于确定汉信码模块矩阵位置的功能图形。

3.5

辅助校正图形 **assistant alignment pattern**

位于汉信码符号边缘,用于同步符号模块的功能图形。

3.6

寻像图形分隔符 **finder pattern separator**

用于将位置探测图形与符号的其余部分分开的功能图形,全部由浅色模块组成的,宽度为一个模块。

3.7

信息编码区域 information encoding region

在符号中没有被功能图形占用,用于对数据或纠错码字进行编码的区域。

3.8

版本 version

用于表示符号规格的数字。

3.9

掩模 masking

为使符号中深色(低反射率)与浅色(高反射率)模块数的比例均衡,在编码区域内对汉信码符号位图的异或(XOR)处理。

3.10

功能信息 function information

用于记录版本、纠错等级、掩模类型的数据位流。

3.11

模式 mode

将特定的字符集表示成二进制位流的方法。

3.12

模式指示符 mode indicator

指示随后的数据序列编码模式的4位标识符。

3.13

模式结束符 mode terminator

用于表示某一编码模式结束的二进制序列。

3.14

二进制计数符 binary counter

定义二进制编码模式下数据串长度的位序列。

3.15

符号填充位 symbol padding bit

当编码区域不能被8位的码字填满时,用于填充最后一个码字后的空位,其值为0,不表示数据。

3.16

扩充解释 Extended Channel Interpretation; ECI

允许对汉信码译码输出数据流与缺省字符集有不同解释的协议。

3.17

信息位流 information bit stream

输入的原始数据经过分模式编码后生成的二进制序列。

3.18

纠错位流 ECC information bit stream

信息位流经纠错编码后生成的用于纠错的二进制序列。

3.19

数据位流 data bit stream

信息位流附加纠错位流后得到的二进制序列。

3.20

填充位 padding bit

用于填充数据位流最后一个码字后面容量的无含义位,其值为0。

4 约定

4.1 数学运算符

本标准中使用的数学运算符定义如下：

Div 整除运算

mod 模运算(求整除后的余数的运算)

XOR 异或运算(当两个输入不不同时,输出为 1;当两个输入相同时,输出为 0)

4.2 模块位置

为便于表述,模块的位置用符号中的行、列坐标表示,格式为 (i, j) 。其中, i 表示模块所在行数(从上向下计算), j 表示模块所在的列数(从左向右计算),记数从 1 开始,因此符号(不包括空白区)中左上角的模块表示为模块(1,1)。

4.3 数据表示形式

$(\dots)_{\text{bin}}$,表示括号中的内容使用二进制表示。

$\dots_{\text{HEX}}$,表示数据内容使用十六进制表示。

不加申明时,1 个字节由 8 个二进制位组成,字节的内容用十六进制表示。

5 符号特性与符号结构

5.1 符号特性

汉信码是矩阵型符号,具有独立定位功能。它还具有如下特性:

5.1.1 基本特性

a) 编码信息

- 1) 数字型字符(数字 0~9);
- 2) 字母型字符(见 GB/T 11383);
- 3) 汉字字符(见 GB 18030);
- 4) 图像、声音等不属于上述类型的信息。

b) 数据表示法

深色模块表示二进制“1”,浅色模块表示二进制“0”。

c) 符号规格

1~84 版本汉信码符号尺寸(不包括空白区)分别为 23×23 模块到 189×189 模块,每一版本符号比前一版本符号每边增加 2 个模块。

d) 编码容量

- 1) 数字:7 827 个字符;
- 2) 字母型字符:4 350 个字符;
- 3) 常用一区汉字:2 174 个字符;
- 4) 常用二区汉字:2 174 个字符;
- 5) 二字节汉字:1 739 个字符;
- 6) 四字节汉字:1 044 个字符;
- 7) 二进制数据:3 261 个字节。

e) 纠错的选择

4 种纠错等级,可恢复的码字比例为:

L1	8%
L2	15%
L3	23%
L4	30%

5.1.2 附加特性

a) 掩模

掩模是汉信码固有的附加特性,可以使汉信码符号中深色与浅色模块的比例接近 1 : 1,使因相邻模块的排列影响高效译码的可能性降为最小。

b) ECI

ECI 是汉信码可选的附加特性,ECI 协议允许对汉信码符号译码输出数据流与缺省字符集有不同解释的协议。这种方式使符号可以表示缺省字符集以外的数据(如阿拉伯字符、古斯拉夫字符、希腊字母等)、其他数据解释(如用一定的压缩方式表示的数据)以及具体行业的特殊编码。

5.2 符号结构

汉信码符号是由 $n \times n$ 个名义正方形模块构成的正方形阵列,该正方形阵列由信息编码区、功能信息区与功能图形区组成,其中功能图形区主要包括寻像图形、寻像图形分隔区、校正图形与辅助校正图形。码图符号的四周为不少于 3 模块宽的空白区。图 1 是 24 版本的汉信码符号结构示意图。

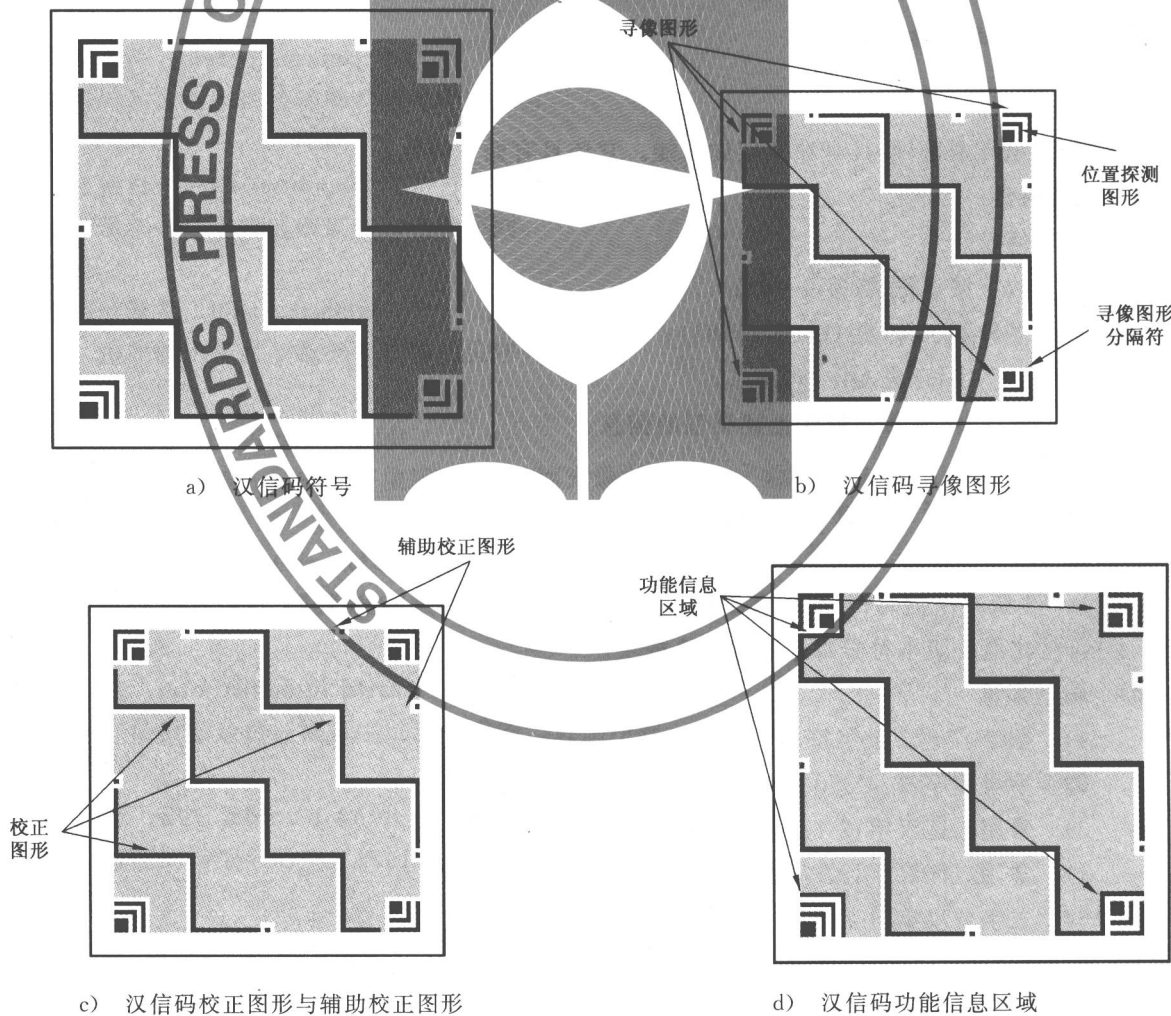
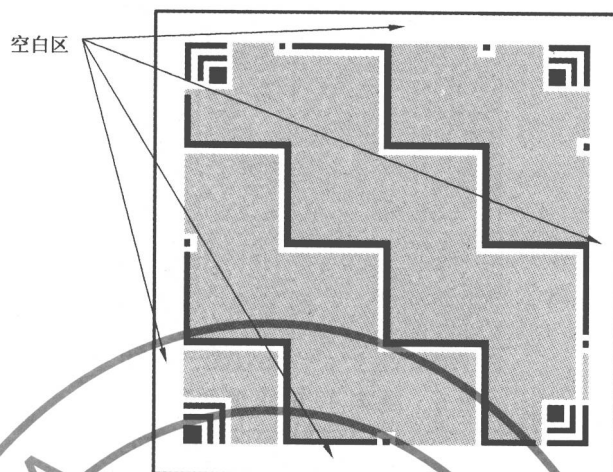


图 1 24 版本汉信码符号结构示意图



e) 空白区

图 1 (续)

5.2.1 符号版本

汉信码符号共有 84 个版本,分别为版本 1、版本 2、……、版本 84。版本 1 的规格为 23 模块×23 模块,版本 2 为 25 模块×25 模块,依此类推,每一版本符号比前一版本每边增加 2 个模块,直到版本 84,其规格为 189 模块×189 模块。图 2~图 7 为版本 1、4、24、40、62 和 84 的符号结构。



图 2 版本 1 的汉信码符号



图 3 版本 4 的汉信码符号

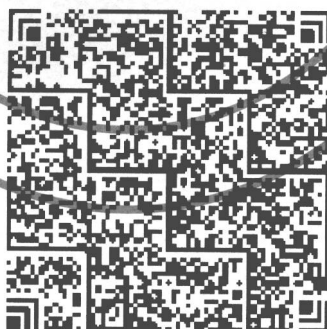


图 4 版本 24 的汉信码符号

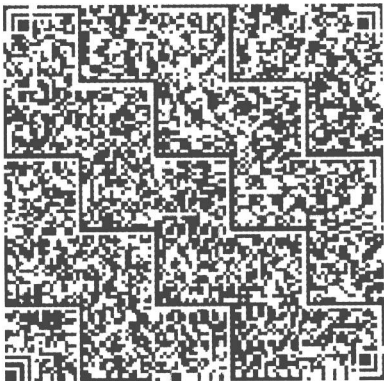


图 5 版本 40 的汉信码符号

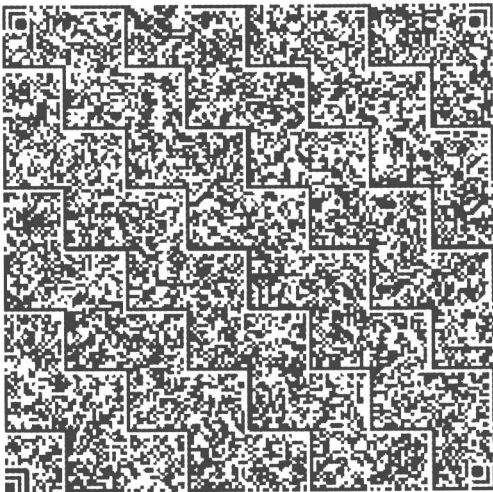


图 6 版本 62 的汉信码符号

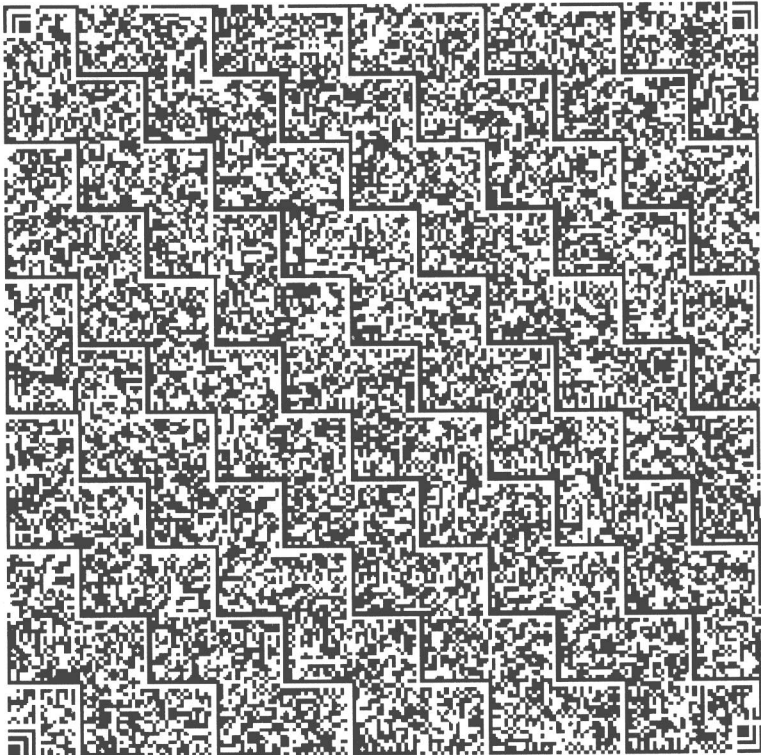


图 7 版本 84 的汉信码符号

5.2.2 寻像图形

汉信码的寻像图形由 4 个位置探测图形组成。分别位于汉信码符号的左上角、右上角、左下角和右下角,见图 1b)。

位置探测图形大小为 7×7 个模块,从结构上,如图 8 所示,可以看作 5 个重叠的依右下角对齐的尺寸不同的正方形组成,它们分别为 7×7 个深色模块、 6×6 个浅色模块、 5×5 个深色模块、 4×4 个浅色模块、 3×3 个深色模块,其扫描的特征比例为 $1:1:1:1:3$ 和 $3:1:1:1:1$ (沿不同方向扫描所得值不同)。位置探测图形中 3×3 深色块的中心称为位置探测中心。符号中其他地方遇到类似比例的可能性极小,因此识别组成寻像图形的四个位置探测图形,可以高速地确定视场中符号的位置和方向。

在位置探测图形的放置上,如图 8 所示的是左上角的位置探测图形,左下角的位置探测图形为左上角的位置探测图形顺时针旋转 90° 得到,右下角的位置探测图形为左下角的位置探测图形顺时针旋转 90° 得到,右上角的位置探测图形为右下角的位置探测图形逆时针旋转 90° 得到。

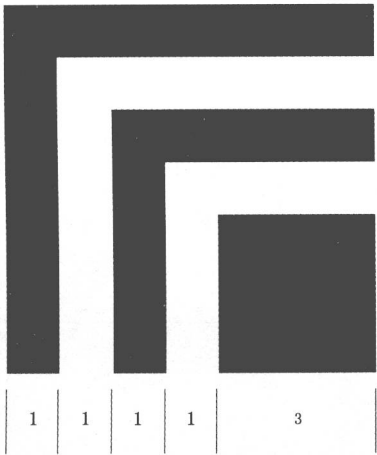


图 8 位置探测图形的结构

5.2.3 寻像图形分隔区

在每个位置探测图形和编码区域之间有宽度为 1 个模块的浅色寻像图形分隔区,如图 1b)所示。

5.2.4 校正图形

汉信码的校正图形是一组阶梯形的校正折线。版本 1、版本 2、版本 3 汉信码符号无校正图形。

校正折线在汉信码符号的左边界和下边界上为 1 模块宽深色条,在符号其他相应位置上由相邻的 1 模块宽的深浅色条组成,遵循深色条在上浅色条在下、深色条在右浅色条在左的原则。如图 9 所示。

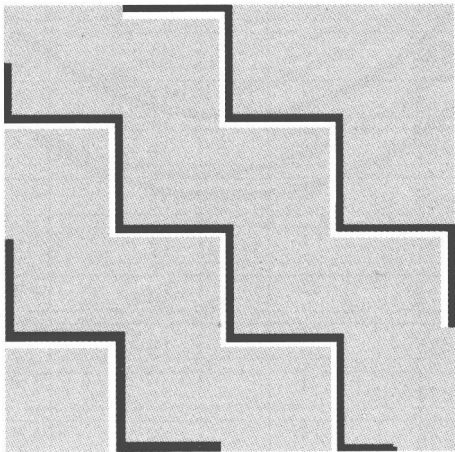


图 9 校正折线

当符号版本为 4~10 时,符号中只有两条校正折线,长度为 k 模块。当符号版本大于 10 时,校正折线的长度分为两种情况:符号左下角的两条校正折线长度是一个特殊值 r 模块;其他区域的校正折线长度相同,为 k 模块。不同版本的符号, r 、 k 、 m 存在下列关系。参数表见表 1。

$$r + m \times k = n$$

式中: n 为汉信码符号单边模块数。

表 1 不同版本符号的校正图形参数表

版本	码图大小(模块)	r (模块)	k (模块)	m
1	23×23	—	—	—
2	25×25	—	—	—
3	27×27	—	—	—
4	29×29	—	14	1
5	31×31	—	16	1
6	33×33	—	16	1
7	35×35	—	17	1
8	37×37	—	18	1
9	39×39	—	19	1
10	41×41	—	20	1
11	43×43	15	14	2
12	45×45	15	15	2
13	47×47	15	16	2
14	49×49	17	16	2
15	51×51	17	17	2
16	53×53	19	17	2
17	55×55	19	18	2
18	57×57	19	19	2
19	59×59	19	20	2
20	61×61	21	20	2
21	63×63	21	21	2
22	65×65	17	16	3
23	67×67	16	17	3
24	69×69	18	17	3
25	71×71	17	18	3
26	73×73	19	18	3
27	75×75	18	19	3
28	77×77	20	19	3
29	79×79	19	20	3
30	81×81	21	20	3

表 1 (续)

版本	码图大小(模块)	r (模块)	k (模块)	m
31	83×83	20	21	3
32	85×85	17	17	4
33	87×87	19	17	4
34	89×89	17	18	4
35	91×91	19	18	4
36	93×93	17	19	4
37	95×95	19	19	4
38	97×97	21	19	4
39	99×99	19	20	4
40	101×101	21	20	4
41	103×103	18	17	5
42	105×105	20	17	5
43	107×107	17	18	5
44	109×109	19	18	5
45	111×111	21	18	5
46	113×113	18	19	5
47	115×115	20	19	5
48	117×117	22	19	5
49	119×119	17	17	6
50	121×121	19	17	6
51	123×123	15	18	6
52	125×125	17	18	6
53	127×127	19	18	6
54	129×129	21	18	6
55	131×131	17	19	6
56	133×133	19	19	6
57	135×135	21	19	6
58	137×137	18	17	7
59	139×139	20	17	7
60	141×141	15	18	7
61	143×143	17	18	7
62	145×145	19	18	7
63	147×147	21	18	7
64	149×149	16	19	7
65	151×151	18	19	7

表 1 (续)

版本	码图大小(模块)	r (模块)	k (模块)	m
66	153×153	17	17	8
67	155×155	19	17	8
68	157×157	21	17	8
69	159×159	15	18	8
70	161×161	17	18	8
71	163×163	19	18	8
72	165×165	21	18	8
73	167×167	15	19	8
74	169×169	17	19	8
75	171×171	18	17	9
76	173×173	20	17	9
77	175×175	22	17	9
78	177×177	15	18	9
79	179×179	17	18	9
80	181×181	19	18	9
81	183×183	21	18	9
82	185×185	23	18	9
83	187×187	17	17	10
84	189×189	19	17	10

5.2.5 辅助校正图形

辅助校正图形为码图 4 个边缘上的 6 个模块(5 个浅色模块,1 个深色模块)组成的图形,如图 10 所示。

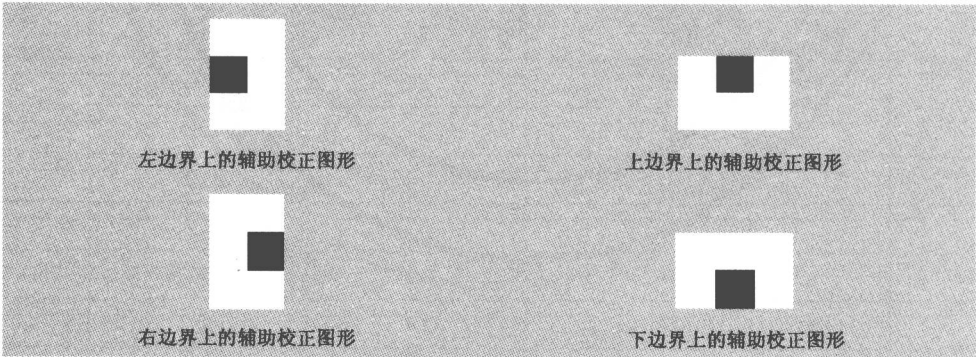


图 10 辅助校正图形

辅助校正图形位于汉信码符号的 4 个边缘上校正折线的端点处以及右上角校正折线延长线与符号边缘的交点处,如图 1c)所示。

5.2.6 功能信息区域

功能信息区域是指符号内围绕 4 个寻像图形分隔区的 1 模块宽的区域,如图 1d)所示。每块功能信息区域由 17 个模块组成,全部功能信息区域包括 68 个模块,用于存放功能信息编码后的数据。

5.2.7 信息编码区

汉信码的信息编码区是符号中除功能图形和功能信息区域之外的区域,用于存放信息编码与纠错编码数据。

5.2.8 空白区

空白区为环绕在汉信码符号四周的不小于 3 模块宽的浅色区域,空白区模块的反射率应与汉信码符号中的浅色模块的反射率相同。

6 数据编码与符号表示

汉信码数据编码与符号表示过程分为数据分析与模式指示、数据编码、纠错编码、最终数据位流构造、码图放置、掩模、功能信息放置七个步骤。

6.1 数据分析与模式指示

6.1.1 数据分析

对需要符号表示的原始数据进行分析,采用缺省的或者其他适当的 ECI 协议,选择适当的编码模式,将原始数据划分为符合一种或多种编码模式的子序列(待编码数据)。汉信码支持 ECI 协议,ECI 协议允许输出的数据流有与缺省的字符集不同的解释。ECI 支持 4 大类解释。

- a) 国际字符集;
- b) 用于诸如加密或压缩等一般目的的解释;
- c) 闭环系统中用户自定义的解释;
- d) 无缓冲模式中用于结构链接的控制信息。

ECI 协议在 AIM ECI 规范中有全面的定义,协议提供了一个对印刷前和译码后的字节值进行详细解释的方法。

汉信码的缺省解释是 ECI 000003 ,表示 GB/T 11383 字符集。

在将原始数据序列划分为子序列的过程中,选择模式时需综合考虑模式转换时产生的模式指示符和结束符等附加数据量。确定了采取的编码模式之后,需要通过所得的信息位流长度和采用的纠错等级,计算预期的数据位流长度,采用与预期的数据位流长度相适应的最小版本汉信码符号。汉信码各版本符号的数据容量见附录 A 。

6.1.2 模式指示

汉信码数据编码模式有数字模式、Text 模式、二进制字节模式、常用汉字 1 区模式、常用汉字 2 区模式、GB 18030 双字节区模式、GB 18030 四字节区模式和 ECI 模式。各种模式由确定的模式指示符指示,模式名称和对应模式指示符见表 2。

表 2 模式指示符

模式名称	模式指示符
数字模式	(0001) _{bin}
Text 模式	(0010) _{bin}
二进制字节模式	(0011) _{bin}
常用汉字 1 区模式	(0100) _{bin}
常用汉字 2 区模式	(0101) _{bin}
GB 18030 双字节区模式	(0110) _{bin}
GB 18030 四字节区模式	(0111) _{bin}
ECI 模式	(1000) _{bin}

汉信码各版本符号与各编码模式的信息容量对应关系见附录 B。

6.1.2.1 数字模式

数字模式对数字型字符(数字 0~9, 字节值 $30_{\text{HEX}} \sim 39_{\text{HEX}}$)编码, 通常对每 3 个数字型字符采用 10 位二进制(即 $(0000000000)_{\text{bin}} \sim (1111100111)_{\text{bin}}$)作为其编码表示, 其指示符为 $(0001)_{\text{bin}}$ 。

6.1.2.2 Text 模式

Text 模式对 GB/T 11383 中规定的常用符号进行编码, 其字节值范围为 $00_{\text{HEX}} \sim 1B_{\text{HEX}}$ 与 $20_{\text{HEX}} \sim 7F_{\text{HEX}}$ 。Text 模式分为两个子模式: Text1 模式和 Text2 模式。其中, Text1 模式编码范围是 GB/T 11383 中规定的大写英文字母(A~Z)、小写英文字母(a~z)和数字(0~9), Text2 模式编码其余的信息数据字符。在 Text 编码模式下每个数据字符用 6 位二进制作为编码表示。Text 模式指示符为 $(0010)_{\text{bin}}$ 。

6.1.2.3 二进制字节模式

二进制字节模式用于表示任意形式的二进制数据, 二进制字节模式采用该二进制数据的字节表示形式作为其编码表示。二进制字节模式指示符为 $(0011)_{\text{bin}}$ 。

6.1.2.4 常用汉字 1 区编码模式

常用汉字 1 区编码模式用 12 位二进制表示, 常用汉字 1 区包括 GB 18030 双字节 2 区中第一字节范围在 $B0_{\text{HEX}} \sim D7_{\text{HEX}}$, 第二字节范围在 $A1_{\text{HEX}} \sim FE_{\text{HEX}}$ (共 3 760 个) 内的汉字和双字节 1 区中第一字节范围在 $A1_{\text{HEX}} \sim A3_{\text{HEX}}$, 第二字节范围在 $A1_{\text{HEX}} \sim FE_{\text{HEX}}$ (共 282 个) 与 $A8A1_{\text{HEX}} \sim A8C0_{\text{HEX}}$ (共 32 个) 的字符, 共 4 074 个。常用汉字 1 区模式指示符为 $(0100)_{\text{bin}}$ 。

6.1.2.5 常用汉字 2 区编码模式

常用汉字 2 区编码模式用 12 位二进制数值表示, 常用汉字 2 区包括 GB 18030 双字节 2 区中第一字节范围在 $D8_{\text{HEX}} \sim F7_{\text{HEX}}$, 第二字节范围在 $A1_{\text{HEX}} \sim FE_{\text{HEX}}$ (即, 从“于”到“龠”) 内的汉字, 共 3 008 个。常用汉字 2 区模式指示符为 $(0101)_{\text{bin}}$ 。

6.1.2.6 GB 18030 双字节区编码模式

GB 18030 双字节区编码模式用 15 位二进制表示 GB 18030 双字节区中所有字符, 即第一字节范围在 $81_{\text{HEX}} \sim FE_{\text{HEX}}$, 第二字节范围在 $40_{\text{HEX}} \sim 7E_{\text{HEX}}$ 和 $80_{\text{HEX}} \sim FE_{\text{HEX}}$ 的字符, 共 23 940 个。GB 18030 双字节区编码模式指示符为 $(0110)_{\text{bin}}$ 。

6.1.2.7 GB 18030 四字节区编码模式

GB 18030 四字节区编码模式用 21 位二进制数值表示 GB 18030 四字节区中所有字符, 即第一字节范围在 $81_{\text{HEX}} \sim FE_{\text{HEX}}$, 第二字节范围在 $30_{\text{HEX}} \sim 39_{\text{HEX}}$, 第三字节范围在 $81_{\text{HEX}} \sim FE_{\text{HEX}}$, 第四字节范围在 $30_{\text{HEX}} \sim 39_{\text{HEX}}$ 的汉字和字符, 共 1 587 600 个。GB 18030 四字节区编码模式指示符为 $(0111)_{\text{bin}}$ 。

6.1.2.8 ECI 模式

本模式用于指明编码数据按照 AIM ECI 规范中定义的 ECI 协议进行解释。ECI 模式用 ECI 模式指示符 $(1000)_{\text{bin}}$ 引入。当开始的 ECI 为缺省 ECI 时, 不需引入 ECI 模式。当发生 ECI 模式转换时, 需要采用 ECI 模式指示符 $(1000)_{\text{bin}}$ 引入 ECI 模式。

ECI 只能用于识读器可以传送符号标识的情况, 不能传送符号标识的识读器无法从包含 ECI 的符号中传输数据。

输入的 ECI 数据需要编码系统作为一系列 8 位字节的值进行处理, 采用数字、Text、常用汉字 1 区等一种或几种模式对其字节值进行最高效编码, 而不必考虑其实际意义。例如: 值为 30_{HEX} 到 39_{HEX} 的数据序列可以当作一个数字(0~9)序列, 用数字模式进行编码(见 6.2.1), 即使实际上它并不表示数字数据。

6.2 数据编码

数据编码分为形成信息位流和构造信息码字序列两个过程。

● 形成信息位流过程

待编码数据应按照数字模式、Text 模式、常用汉字 1 区等模式进行编码, 各模式编码过程见 6.2.1~

6.2.8。每个模式编码形成的位流由以下内容组成: