

# 一种屏幕图文高效压缩存储系统的实现

方 斌

(安徽大学电子工程与信息科学系 合肥 230039)

**摘要** 本文针对屏幕图文存储量不足问题,提出了一种新的压缩存储系统的原理和设计。新系统具有高压缩比和快运行速度。

**关键词** 图文,压缩存储,代码

## 1 引言

随着微机图形技术的发展,计算机屏幕图文信息存储越来越广泛地被使用。一般高级语言都向用户提供了屏幕图文存储的功能。若对图文信息不经压缩处理,直接存储,就存在占有存储容量大和速度慢等缺点。对  $640 \times 200$  图形显示方式,需 16kB 存储空间。为提高计算机存储空间的利用率,需采用压缩存储技术,间接地减小屏幕数据占用的容量。数据压缩一般分为无失真压缩和有失真压缩。本文介绍一种简便的无失真压缩系统的设计。

## 2 压缩思想

由于在图形工作方式下,进行屏幕任意的文字显示和作图,其对应内存映象是一组杂乱无规律的二进制数据。对这样的数据,必须找出一些共同特性,作为压缩存储依据。

事实上屏幕图文中存在大量空白统计,这为压缩存储提供了有效处理的思路。屏幕上空白对应内存映象是零字节。如仅存非零的字节,则可大幅度减少存储容量。第一步一个有效的方法是顺序处理屏幕映象数据,将零元素剔除并记忆连续为零字节的个数  $N_0$ 。由于对映象数据进行顺序处理,因而无需记忆零字节所在位置,但在恢复时要区别零字节个数  $N_0$  和非零字节,为此需在个数  $N_0$  前加一识别代码。第一步处理后的数据如图 1 所示顺序表示。代码 1 表示该位置为零字节并且其后有  $N_0$  个连续的零字节。

当  $N_0$  为 1 或 2 出现次数较多时,若仍按图 1 结构进行压缩就达不到应有效果。个数  $N_0$  为 1 时用代码 1 和个数  $N_0$  两个字节表示一个零字节反而会增加数量,因此就有必要在剔除零的同时作另一处理,即将屏幕映象数据中零字节连续个数  $N_0$  为 1 时,仍存一个零字节;同样  $N_0$  为 2 时用两个字节表示两个零字节就没有起到压缩作用,这时可用另一识别代码 2 表示该位置数据为零且连续个数为 2; 经过这步处理就可避免压缩处理的失效。该步处理后的数据按图 2 所示顺序表示。其中代码 2 代表两个零字节,零字节是指一个零字节。

通过大量观察分析,有很多非零的数据与其前 80 位、8 位和 16 位数据相同,并且有

1994-02-14 收到,1994-07-01 定稿

方 斌 男,1966 年生,讲师,现从事 CAD 和智能控制方面的教学和研究工作。

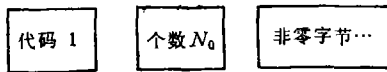


图1 第一步压缩的数据结构

图2  $N_0$  为 1,2 的数据结构

很多连续相同的情况。这反映在屏幕上相邻行之间某段具有相同的图形。第二步就是将非零字节与前 80 位、8 位和 16 位数据进行比较,若相同就用识别代码加连续相同的字节个数  $N_i$  代替原有的  $N_i$  个非零数据。同样原因,若相同个数  $N_i$  为 1 则仍存原数据。代码 3,4,5 分别表示与前 80 位、8 位、16 位数据相同的情况,其数据结构如图 3,其中  $N_i$  表示连续相同的字节个数。

第一步压缩实际上是进行列的压缩,第二步进行了行的压缩。经过上述处理后的  $N$  个数据,仍可发现一些特点,即有很多非零字节的前 4 位或后 4 位为零;这就为下一步压缩提供了依据。这步处理是在前两步处理后的  $N$  个数据上进行,将每一字节分为前 4 位和后 4 位两个数,两数中若有为零的就剔除;依次处理每一字节,用两个不为零的 4 位按先后顺序组成新的字节。为识别哪 4 位为零,将最后存储的数据段前  $M$  个字节 ( $M = \text{INT}(N/4)$ ) 作识别符。 $M$  个识别符后为按先后顺序存储的新字节。前  $M$  个二进制中每两位表示压缩前的一个字节,偶数位表示压缩前字节的前 4 位,奇数位为压缩前的后 4 位。前  $M$  个二进制中的 1 表示所对应的 4 位非零,可以新产生的字节按顺序读取 4 位;0 表示所对应的 4 位为零。

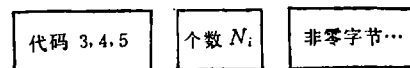


图3 第二步压缩的数据结构

经过前三步压缩后的数据结构如表 1 所示:

表 1 压缩后的数据结构

字节 1	...	字节 5	字节 6,7	字节 8...字节 7+M	字节 8+M...
代码 1	...	代码 5	数 M	识别符	压缩后数据

### 3 压缩系统的实现

为便于高级语言的调用和提高运行速度,选用汇编语言进行编程。压缩子系统见框图 4。恢复子系统为压缩处理逆过程。首先将表 1 所示的数据结构恢复成图 1 到图 3 所示的数据结构,然后根据 5 种代码恢复相应数据直接写入屏幕内存。为保证无失真的压缩,首先必须产生 5 个代码并保证代码与图文数据不相同;若有相同就会造成非零字节中某个字节起到不应是代码的作用,使该位置产生失真;并造成其后的数据向后移动。经大量统计分析,屏幕内存映象中很少出现  $iiH$  数据 ( $i = 1, 2, \dots, E$ ),即使有,在这 14 种数据中某个别数据仅能出现 1 到 2 次。从屏幕上看这些数所对应的象素较为奇异,如 AAH 对应象素为 10101010。首先对屏幕内存映象预处理一遍在这 14 个数中产生 5 个代码。若不能生成 5 个代码,还可从  $iA, iB$  等 ( $i = 1, 2, \dots, E$ ) 数据产生。

### 4 应用结论

本系统作为题库的存储子系统已通过专家鉴定。经过几年的实际应用,具有较好的

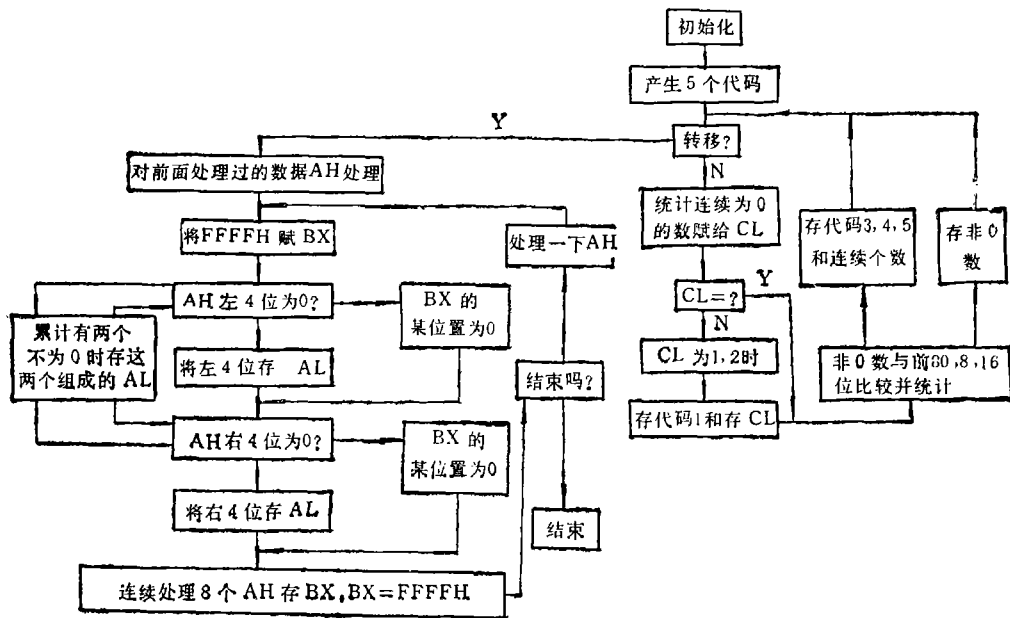


图 4 压缩系统程序框图

CL: 统计连续为 0 的数赋给 CL. AH: 前两步压缩后数据送入 AH.  
AL: 连续两个非零 4 位组成一个字, BX: 识别符, 初始置为 FFFFH.

压缩效果, 且不因图形复杂而出现相反的效果。一般情况下屏幕一半为汉字一半为图形时经压缩处理后容量仅近 2kB, 压缩比平均为 8 左右。即使整个屏幕写满各种汉字, 压缩后也仅为 5kB; 且不会出现失真。同时压缩或恢复系统运行速度快, 仅数秒内便可完成压缩或恢复。

## 参 考 文 献

- [1] 沈美明, 等. IBM PC 汇编语言程序设计, 北京: 清华大学出版社, 1987, 第十章.
- [2] 杨钟藩. 微型计算机绘图及其程序设计. 北京: 电子工业出版社, 1987, 64—110.
- [3] 钱培德. 微型计算机汉字操作系统. 西安: 陕西电子编辑部, 1987, 第八章.

## IMPLEMENTATION OF AN EFFICIENT COMPRESSION STORAGE SYSTEM FOR SCREEN GRAPHICS-LETTER

Fang Bin

*(Department of Electronic Engineering and Information Science, Anhui University, Hefei 230039)*

**Abstract** This paper discusses the principle and design of the implementation of a compression storage system for screen graphics-letter. The new system overcomes the shortage of screen storage, and possesses a large compression ratio and a high operating speed.

**Key words** Graphics-letter, Compression storage, Code