

一种特殊图像数据压缩存储及差错控制方法

颜昌翔, 于平, 王家骥

(中国科学院长春光学精密机械与物理研究所, 吉林 长春 130022)

摘要: 针对工程中特定(尺寸固定、冗余度较高)图像数据的无损压缩存储问题, 提出了一种基于行程的压缩编码方法以及简单的差错控制编码方法。简要介绍了问题的特点, 探讨了应用几种图像压缩及容错编码方法的可能性; 通过选择特定方法(类似 JPEG 的压缩方法和简单的二维奇偶校验编码)和进行具体地压缩编程实践, 找到了较优的方法; 最后对此方法作了简单分析并给出了结果。该方法具有效率较高、存取迅速、简单实用和可靠度高等特点。

关键词: 无损压缩; 差错控制; 行程编码

中图分类号: TP391.4 **文献标识码:** A

1 引言

为了存储、处理和传输大量数据, 有时必须进行压缩和差错控制。数据压缩方法种类繁多, 可以分为无损压缩和有损压缩两大类^[1]。有损压缩可以得到很高的压缩率, 但会引入失真。无损压缩利用数据的统计冗余进行压缩, 可完全恢复原始数据而不引入任何失真, 压缩率一般为 2:1 ~ 5:1。这类方法广泛用于文本数据、程序和特殊应用场合的图像数据(如指纹图像、医学图像等)^[1-3]。

差错控制编码技术是适应数字信息抗噪声干扰需要而诞生和发展起来的。如果在一个字中具有永久性的错误, 则称为硬错误(hard error); 如果错误只有暂时的性质(例如由于高能粒子对半导体存储器的某一比特位的干扰), 则称为软错误(soft error)。硬错误包括存储单元错、存储芯片内的行错或列错、存储芯片全失效等。根据统计, 存储单元失效约占芯片失效总数的 35%。对 16k 动态 RAM 的观测表明, 软故障与硬故障之比为 4:1。一位错误占全部错误率的 90% 以上^[4]。

本文要讨论的是图像数据压缩和差错控制问题: 灰级图像大小约为 3200 × 1000, 其特点是 80% 左右的数据是 0, 而所有像素灰度值都介于 0

~ 200, 根据工程实际要求压缩比为 6:1, 数据不失真, 采用与实际有效数据同样大小的数据空间进行数据差错控制编码, 达到纠 1 位错误的水平。同时要求可由 16 位处理器进行实时随机读取。

2 压缩和存储方案

2.1 分块行程无损压缩方案^[1,5]

静态图像压缩方法, 目前较新颖的主要有小波变换方法和分形方法, 但计算量都较大, 不适用于本问题。JPEG 是静态图像的第二个国际数字图像压缩标准, 也是至今一直在使用的、应用最广的图像压缩标准, 其压缩模式有以下几种:

顺序式编码(Sequential Encoding);

递增式编码(Progressive Encoding);

无失真编码(Lossless Encoding);

阶梯式编码(Hierarchical Encoding)。

采用两种编码机制: 一是行程长度编码; 二是熵编码(Entropy Coding)。

借鉴 JPEG 的编码特点, 采用分块后顺序式无失真编码模式及行程长度编码机制, 较适合于本文中的图像压缩问题。同时应通过一定的手段, 充分利用数据码的特点, 来提高编码效率。

2.2 二维奇偶监督差错控制方案^[4]

为解决通讯和存储等的数据可靠性, 编码学者发明了大量的容错编码方法。容错编码的原理是较小的错误量占全部错误量的绝大部分比例, 这样就可以通过对数据进行适当的编码, 使小量错误得到纠正, 并对大部分的较大量错误得到报警。但出于实时随机读取数据的要求, 应采用计算量较小的方案。

奇偶校验码是一种最简单的检错码, 也是计算机系统最常用到的一种差错控制码。它的编码方式是在 $n - 1$ 位信息元 ($C_{n-1}C_{n-2} \dots C_1$) 后附加 1 位监督元 C_0 , 使得码字中 1 的数目保持奇数或偶数。以构成奇监督或偶监督。

为了克服一维奇偶监督无法检测出两位或偶数位错误的问题, 除了水平方向奇偶监督外, 人们进一步考虑在垂直方向或斜对角线方向上增加一维奇偶监督, 即构成二维水平——垂直奇偶监督码或二维水平——斜向奇偶监督码。

二维奇偶监督码所用的监督元数比一维奇偶监督码要多, 但它能纠正一位错误, 检出绝大部分错误。如果按行向(或列向)依次发送, 则可以检出突发长度 $< m$ (或 n) 的所有突发错误。因此, 在一些计算机网络数据传输系统, 例如 RDC 网的基本数据通信协议 BL 和高级数据通信协议 HL 中, 就曾分别采用上述两种二维奇偶监督码。

本文中准备采用一倍存储空间作为备份空间, 此空间足以用来存放对原始数据的二维奇偶监督码。为减少运算时间, 可以尽量利用此备份空间, 降低编码复杂度。

3 编码方法

3.1 压缩算法

本方案压缩的原理是行程方法, 即一系列连续的相同的数用两个数表示(称合并), 其中前一个数代表其长度, 后一个数代表其大小。也就是 $\boxed{\text{SIZE}} + \boxed{\text{VALUE}}$ 的形式(称为合序列)。这种方法的特点是查找时要经过多次读取和计算, 而且整个数列具有整体性, 若中间发生一个 $\boxed{\text{SIZE}}$ 错误, 则数列后面部分无法读取。为了降低风险, 缩短查找时间, 应将全部数据分块存储, 而且应使数据块足够小。

分块的方法, 应便于用简单的算法生成各块的索引地址。通过索引地址找到各数据块的内容所在空间区的地址。在每个数据块的存储空间内,

数据将按 $\boxed{\text{SIZE}} + \boxed{\text{VALUE}}$ 的形式依次存放, 其中有一些 $\boxed{\text{SIZE}}$ 为 1 的序列。当很多这样的序列放在一起时, 将会带来空间的浪费。故采用一种 $\boxed{\text{SIZES}} + \boxed{\text{VALUES}}$ 的方法(称为独序列), $\boxed{\text{VALUES}}$ 代表一个序列(利用图像分块大小限制, 通过一定措施, 这里 $\boxed{\text{SIZES}}$ 可以与 $\boxed{\text{SIZE}}$ 相区分), 实际长度是 $\boxed{\text{SIZES}}$, 而值的大小分别是其中的各个值。

因为 16 位处理器的分段寻址方式中, 段内地址定义为寻址 64K 字节。压缩图像只能分段存放, 每段大小为 64K 字节。每段分前后两部分, 前面部分是依次存放的段内各个数据块的索引数据, 后面部分是依次存放的由索引所指向的图像块的压缩数据。每条索引由三字节组成: 前两字节的数据块的首地址, 后一字节的数据块特征标志。

压缩步骤:

a) 数据块划分: 全部数据分为 10×10 大小的数据块, 并按先横向后纵向的顺序排列。

b) 数据块数据序列的生成: 将 10×10 的数据按先横向后纵向的顺序排列成一序列。

c) 数据块数据压缩序列的生成: 将 100 个数的序列表示为合序列和独序列依次排列的形式。压缩的规则:

(1) $\boxed{\text{SIZE}}$ 大于 3, 合并;

(2) $\boxed{\text{SIZE}}$ 等于 3, 均为独序列则不合并(前或后无序列或相邻有合序列则合并);

(3) $\boxed{\text{SIZE}}$ 等于 2, 相邻均为合序列则合并(前或后无序列, 则不合并);

(4) 若总长度大于或等于 100, 或分序列数大于 30 个, 则不进行任何合并。

d) 数据块索引的生成: 生成各个数据块的索引, 每个索引用 $\boxed{\text{ADDRESS}} + \boxed{\text{NUM}}$ 的形式, 共三字节。 $\boxed{\text{ADDRESS}}$ 是数据块对应的地址, 为两字节; $\boxed{\text{NUM}}$ 是数据块的标志, 为 1 字节, 代表数据块压缩序列中包含序列的数目。

e) 数据块索引分段数组的生成: 每段空间中的数据块数目不定, 故而要在生成数据块索引的过程中把每段空间的数据块数目用数组的形式记录下来。生成分段数组的过程是逐步试探中完成的。

3.2 差错控制算法

本文采取的容错编码方法是采用 4 位信息元和 4 位监督元, 计算关系如下:

$S_1S_2S_3S_4R_1R_2R_3R_4$, S_i 代表信息元, R_i 代表监督元。

编码计算: $R_1= S_1+ S_2$ $R_2= S_3+ S_4$ $R_3= S_1+ S_3$ $R_4= S_2+ S_4$

译码计算:

a) $P_1= R_1+ S_1+ S_2$ $P_2= R_2+ S_3+ S_4$ $P_3= R_3+ S_1+ S_3$ $P_4= R_4+ S_2+ S_4$;

b) 若 $P_1 \sim P_4$ 全为 0 或有 3 个为 0, 输出 $S_1S_2S_3S_4$;

c) 若 $P_1 \sim P_4$ 有两个 1,

(1) P_1P_3 为 1, 输出 $\overline{S_1}S_2S_3S_4$;

(2) P_1P_4 为 1, 输出 $S_1\overline{S_2}S_3S_4$;

(3) P_2P_3 为 1, 输出 $S_1S_2\overline{S_3}S_4$;

(4) P_2P_4 为 1, 输出 $S_1S_2S_3\overline{S_4}$;

(5) P_1P_2 或 P_3P_4 为 1, 判断有两个错。

d) 若 $P_1 \sim P_4$ 有三个 1, 判断有 2 个错。

这样, 就可以进行纠正 1 位错误, 并且能判断约 90% 的 2 位错误。

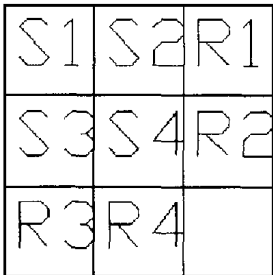
4 编码结果

4.1 压缩编码

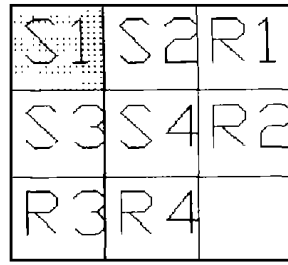
根据算法, 每分块的序列经过处理后, 最多是由 30 个分序列组成, 即随机查找某一个像素, 最多需要进行 30 次左右的存储器读操作。通过实验, 平均读操作次数不超过 4.5 次。

4.2 差错控制编码

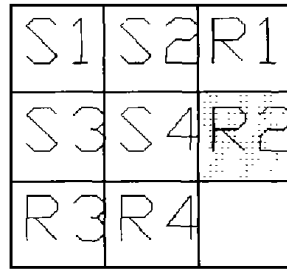
采用该算法, 充分利用了备份空间, 减少了编码和解码时间, 为实时随机读取图像数据提供了条件, 同时又完成了纠 1 位错误和基本判断 2 位错误的水平。



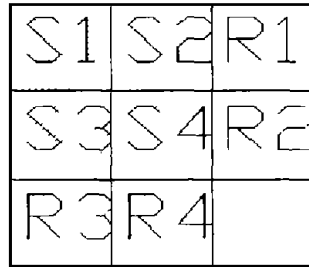
a. No error



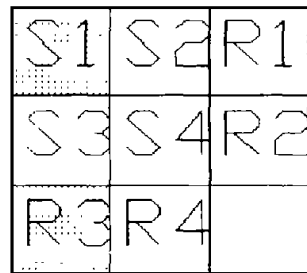
b. one-bit info error



c. One-bit monitor error



d. Two-bit error can be decided



e. two-bit error can not be decided

Fig. 1 Some situations for error control

如图 1 所示, 其中 a 表示无差错情况, b 表示有一位信息元错误, c 表示有一位监督元错误, d 表示有二位信息元错误, e 表示有一位信息元和一位监督元错误并且不能纠错的情况。

5 结 论

在综合各类压缩算法后,根据实际图像特点,对最常见的 JPEG 压缩算法作一些改动,使算法充分利用到图像特点,达到了压缩比要求;根据差

错控制编码理论,选取了最简单的二维奇偶校验的差错控制编码方法,达到了纠 1 判 2 的一般要求,提高了图像数据的可靠性;编码算法充分考虑了 16 位微处理器的实时随机读取图像像素的要求,使计算量达到了最小。该方法具有效率较高、存取迅速、简单实用和可靠度高等特点。

参考文献:

- [1] 吴乐南. 数据压缩的原理与应用[M]. 北京: 电子工业出版社, 1995.
- [2] 张雪松, 等. 图像编码技术发展综述[J]. 光学技术, 1997, (3): 39- 41.
- [3] 贾晓光, 等. 航天遥感图像压缩技术的最新发展[J]. 中国图像图形学报, 1997, 2(10): 697- 700.
- [4] 靳蕃, 陈志. 组合编码原理及应用[M]. 上海: 上海科学技术出版社, 1995.
- [5] 闫丽, 袁家虎, 刘江, 李晨. 光学子波用于图像处理[J]. 光学 精密工程, 2000, (8): 225- 230.

Method for compression and error control of a special picture

YAN Chang-xiang , YU Ping , WANG Jia-ti

(*Space Optics Department, Changchun Institute of Optics,
Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130022, China*)

Abstract: Aiming at the problems in lossless compression and storage of an image in the actual work, this paper presents a compression method based on run-length lossless encoding and an error control coding method applied to the special image in fixed dimension and with high redundancy. Features of the problem were introduced briefly. The possibility of applying several image compressing and coding methods to the problem was discussed. A special compression method, similar to JPEG compression coding and a simple two dimensions odd-even error control method, was applied to the coding practice. This method was analyzed simply and the coding result was given. The method presented in this paper has merits as: high efficiency, rapid access, simplicity and practicality, and high reliability.

Key words: lossless compression, error control, run-length coding

作者简介: 颜昌翔(1973-), 男, 湖北省洪湖市人, 1995年毕业于长春光学精密机械学院光学工艺及测试专业, 获工学学士。1998年毕业于浙江大学光学仪器专业, 获工学硕士。现为中科院长春光学精密机械与物理研究所机械制造及自动化专业博士研究生。