

二元序列游长多次缩减的无损压缩编码方法

高 健, 饶 珺, 孙瑞鹏

(上海大学 机电工程与自动化学院, 上海 200072)

摘 要:文中提出了一种二元序列游长多次缩减的无损压缩编码方法,该方法可依二元序列中游程的分布特征,选取游长缩减起始值。对二元序列中游程长度大于或等于游长起始缩减值的游程进行长度缩减,而对小于游长起始缩减值的游程不做缩减处理,避免了对较短游程重新编码所导致的比特数增加,从而实现二元序列中较长游程的压缩游长缩减。本方法可在二元序列中反复进行直至二元序列不可进一步压缩为止。通过与国际传真标准 CCITT. 4 (G3) 进行试验对比,证明了文中提出的方法对二元序列的压缩具有更高的压缩比和编码效率。

关键词:二元序列;游长缩减;游长缩减起始值;无损压缩

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2013)06-0031-04

doi:10.3969/j.issn.1673-629X.2013.06.008

A Lossless Compression Algorithm of Multiple Run-length Reduction for Binary Sequences

GAO Jian, RAO Jun, SUN Rui-peng

(College of Mechatronic Engineering and Automation of Shanghai University, Shanghai 200072, China)

Abstract: This research proposes a lossless compression algorithm of multiple run-length reduction for binary sequences, which depends on the characteristics of run-length, an initial-value can be chosen in this algorithm. This research only reduces the run-lengths which are greater or be equal to the initial-value, while for the purpose of decreasing bit number, the other run-lengths are reserved, then the greater run-lengths have been reduced. This compression algorithm of run-length can be repeated until the binary sequence can't be compressed. Compared with CCITT. 4 (G3) by lager amount of experiments, the results show that this algorithm has a higher compression ratio and code efficiency for binary sequences.

Key words: binary sequence; run-length reduction; the initial value of reduction; lossless compression

0 引 言

在数字图像中,相邻像素之间存在空间冗余度,图像压缩^[1,2]主要是通过减少冗余来达到减少存储空间的目的。图像压缩按照压缩过程中是否有信息的损失以及解压后与原始图像是否有误差可以分为无损压缩和有损压缩两大类^[3]。在数字信息领域中,常可将页面文字、工程图纸、二值图像等归纳为一个二元序列。在一个二元序列中,相同元素组成的序列段称为一个游程,长度称为游程长度,简称游长。游程编码^[4,5](run-length coding)也叫行程编码,它是用于二值数据的一种应用比较广泛的编码方法,原理是将游程用一个计数值(游程长度)和元素值来代替,根据二元序列

中游程长度的分布特征,基于变长编码原理给出相应游程长度的编码,从而实现二元序列的无损压缩。

国际传真标准 CCITT. 4 (G3)^[6]采用的是一维游程编码,利用霍夫曼方法^[7]对游程长度进行编码。CCITT. 4 (G3)标准根据游程长度大于或小于等于63,将码字分为形成码和终止码两种,对于0~63之间的游长,用单个码字,即终止码表示;对于大于63的游长,通过形成码和终止码的组合来表示。CCITT. 4 (G3)标准的解码过程是通过编码表逐一匹配并恢复各游程长度。上述传统的游程编码方法,对二元序列中全部游程重新进行编码^[8],往往导致较短游程的编码位数大于较短游程长度的自然位数,当二元序列中

收稿日期:2012-09-04

修回日期:2012-12-08

网络出版时间:2013-05-14

基金项目:国家自然科学基金资助项目(60874076)

作者简介:高 健(1961-),男,湖北武汉人,副教授,主要研究方向为图像处理、计算数学;饶 珺(1990-),女,江西抚州人,硕士研究生,主要研究方向为数字图像处理。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20130514.1707.006.html>

较短游程较多时,较短游程重新编码所导致的数据膨胀会严重影响二元序列的压缩效能^[9]。

针对上述问题,文中提出了一种具有自适应特性的游程编码无损压缩算法,依据二元序列中游程的分布特征,选择一个游长缩减起始值,对二元序列中长度大于或等于游长缩减起始值的游程进行缩减,而对小于游长缩减起始值的游程不进行缩减处理(不重新编码),从而实现在对二元序列中较长游程缩减的同时,避免了因对较短游程重新编码所导致的数据膨胀^[10,11]。此外,文中提出的方法可在二元序列中进行多次游长缩减直至达到最佳压缩效果。

1 游长多次缩减编码方法

1.1 基于游长缩减起始值的游长缩减

为论述方便,将二元序列中二元的取值规定为“0”和“1”,并将由 0 元素或 1 元素组成的子序列称为 0 游程或 1 游程,对应的游程长度分别称为 0 游长或 1 游长。基于 0 游程和 1 游程是交替方式出现的,故可通过二元序列中第一个游程属性(0 游程或 1 游程)以及后续的游程长度,表述任意的二元序列。设 B 为一个二元取值为 0 和 1 的二元序列,将二元序列 B 转换为游长序列,记为:

$$L = \{l_0, l_1, l_2, \dots\} \quad (1)$$

设游长缩减起始值为 l ,对游长序列 L 的游长缩减按下述公式进行:

$$\begin{cases} \bar{l}_i = \text{round}(l_i + D/2) & l_i \geq l \\ \bar{l}_i = l_i & l_i < l \end{cases} \quad (2)$$

其中 round 为取整函数,游长序列 $L = \{l_0, l_1, l_2, \dots\}$ 中游长大于等于 l 的全体游长 l_i 加上 l 除 2 的余数依次计入序列 R ,序列 R 为游长序列转换关系码表。

经游长缩减后,原始游长序列 $L = \{l_0, l_1, l_2, \dots\}$ 转换为:

$$\bar{L} = \{\bar{l}_0, \bar{l}_1, \bar{l}_2, \dots\} \quad (3)$$

对游长序列 \bar{L} ,将其中游长大于或等于 l 的游长 \bar{l}_i 乘 2 减去 l ,加上转换关系码表 R 中相应的转换关系码元(0 或 1),即可复原原始游长序列 L 。

1.2 游长缩减起始值的确定

游长缩减起始值是游长缩减方法的一个关键的参数,文中采用试探方法从 $1, 2, \dots, r$ ($r=8$ 或 16) 中选取编码位数最小时所对应的值,作为游长缩减起始值,具体如下:

一般而言,二元序列在不同区域的分布是不同的,为使游长缩减方法达到更好的压缩效果,文中将游长序列 $L = \{l_1, l_2, \dots, l_n\}$ 按每组 1024 个游长单元对游长

序列 L 进行分组,即有:

$$L = \{L_1, L_2, \dots, L_j, \dots, L_N\} \quad (4)$$

其中,游长子序列 $L_j = \{l_{j0}, l_{j1}, \dots, l_{j1023}\}$, $N = \text{round}(n/1024)$,最后一组游长子序列 L_N 元素个数为 $1024 + n \% 1024$,“ $\%$ ”表示求余。

按游长子序列 L_j 中各游长下标的奇偶性,将游长子序列 L_j 分为下述两个子序列

$$L_{je} = \{l_{j0}, l_{j2}, \dots, l_{j1022}\}, L_{jo} = \{l_{j1}, l_{j3}, \dots, l_{j1023}\} \quad (5)$$

分别称 L_{je} 和 L_{jo} 为游长子序列 L_j 中的偶子序列和奇子序列。

统计偶子序列 L_{je} 的各个游长分布个数,得到偶子序列 L_{je} 的游长分布直方图 h_{je} 。选取参数 $m, 1 \leq m \leq r$ ($r=8$ 或 16),对偶子序列游长分布直方图 h_{je} 中各非零元素所对应的游长按公式(2)进行游长缩减。如果游长小于 m ,则将原游长乘上 h_{je} 中对应的游长个数,由此得到偶子序列 L_{je} 中游长小于 m 的全体游长编码位数 $B_{je1}(m)$;如果游长大于或等于 m ,则将游长缩减值加 1(此处“1”表示转换关系的一位码元),并乘上游长分布直方图 h_{je} 中相应游长元素的统计个数,得到偶子序列 L_{je} 中游长大于或等于 m 的全体游长编码位数 $B_{je2}(m)$ 。称 $B_{je}(m) = B_{je1}(m) + B_{je2}(m)$ 为偶子序列 L_{je} 在参数 m 下的编码位数。

令:

$$B_{je}(p_j) = \min\{B_{je}(1), B_{je}(2), \dots, B_{je}(n)\} \quad (6)$$

且当:

$$B_{je}(p_j) < l_{j0} + l_{j2} + \dots + l_{j1022} \quad (7)$$

称 p_j 为偶子序列 L_{je} 的游长缩减起始值。如果式(7)不成立,表明偶子序列 L_{je} 的游长缩减值不存在,意味着偶子序列 L_{je} 不能通过游长缩减的方式进行游长缩减。按上述相同方式,可获得奇子序列 L_{jo} 的游长缩减起始值 q_j 。

1.3 游长的多次缩减

根据上述讨论,若式(7)成立,则可对游长子序列中偶游长子序列或奇游长子序列基于相应的游长缩减起始值进行游长缩减,游长缩减过程按公式(2)进行。对于较长游长,通过 1 次游长缩减后,其游长大约缩减至其原始游长的一半,仍具进一步缩减的可能。为此,文中采用游长多次缩减的方法对游长子序列进行不断的缩减,直至达到最佳压缩效果,具体步骤如下:

(1) 最大缩减次数确定。

分别计算游长序列 $L = \{l_0, l_1, l_2, \dots\}$ 中奇偶游长的算术平均值 $l(o)$ 和 $l(e)$,当:

$l(\cdot) < 16$,则确定奇游长或偶游长子序列的最大缩减次数为 8;

$l(\cdot) \geq 16$,则确定奇游长或偶游长子序列的最大缩减次数为 16。

当最大缩减次数为8或16时,则1.2中所述参数 $m(1 \leq m \leq r)$ 的最大取值 r 为8或16。

(2) 游长子序列 L_j 的多次缩减。

如果 $l(o) < 16$, 则从1,2,...,8中,依游长缩减起始值的确定方法,选取游长子序列 L_j 的奇游长子序列 L_{jo} 的游长缩减起始值 q_j ; 如果 $l(o) \geq 16$, 则从1,2,...,16中,依游长缩减起始值的确定方法,选取游长子序列 L_j 的奇游长子序列 L_{jo} 的游长缩减起始值 q_j 。按相同方式确定偶游长子序列的游长起始值 p_j 。

按照公式(2)所述方法,分别对奇偶游长子序列基于相应的游长缩减起始值进行游长缩减,余数依次计入 R_j 中, R_j 称为游长子序列 L_j 的转换关系码表。通过游长缩减后,奇偶游长子序列 L_{jo} 、 L_{je} 分别转换为 $L_{jo}(1)$ 和 $L_{je}(1)$ 。

对经过游长1次缩减后的奇偶游长子序列 $L_{jo}(1)$ 和 $L_{je}(1)$, 重新选取奇偶游长子序列的游长缩减起始值,并按公式(2)所述方法进行第2次游长缩减,奇偶游长子序列 $L_{jo}(1)$ 和 $L_{je}(1)$ 分别转换为 $L_{jo}(2)$ 和 $L_{je}(2)$, 第2次缩减过程中产生的余数依次计入 R_j 中。以此类推,按奇偶游长的最大缩减次数,分别对奇偶游长子序列 L_{jo} 、 L_{je} 分别进行多次游长缩减直至生成 $L_{jo}(t)$ 和 $L_{je}(s)$ 。

为了方便恢复原始游长子序列 L_j , 将转换关系码表 R_j 进行倒序置放,即首先在 R_j 中置入最后1次缩减的余数码,接着在 R_j 中依次置入倒数第2次缩减的余数码,最后在 R_j 中置入第1次缩减的余数码。

(3) 游长子序列 L_j 多次缩减后游长编码表。

游长子序列 L_j 多次缩减后游长编码表由5部分组成,参照表1,第1部分中的 S_e 和 S_o 为游长子序列中偶游长和奇游长的压缩状态标记(0或1),0表示对应的奇游长子序列或偶游长子序列不可缩减,1表示可缩减;第2部分中的 s 和 t 为偶游长子序列与奇游长子序列的缩减次数,如果偶游长最大缩减次数小于16,则 s 的位数为3位,否则 s 的位数为4位, t 的位数按相同方式确定;第3部分 e_1, e_2, \dots 和 o_1, o_2, \dots 分别为偶游长子序列和奇游长子序列的各次缩减起始值,如果 s 的位数为3位,则 e_1, e_2, \dots 的位数为3位,否则 e_1, e_2, \dots 的位数为4位, o_1, o_2, \dots 的位数按相同方式确定;第4部分为游长子序列 L_j 经多次缩减后生成的游长子序列 $L_j(s|t)$ 所对应的二值码子序列。 $L_j(s|t)$ 表示游长子序列中偶游长子序列经 s 次缩减,奇游长子序列经 t 次缩减后的游长子序列。

如果游长编码表中第1部分码元 $S_e = 0, S_o = 1$, 则游长编码表中第2部分码元为 t , 第3部分码元为 o_1, o_2, \dots ; 如果游长编码表中第1部分码元 $S_e = 1, S_o = 0$, 则游长编码表中第2部分码元为 s , 第3部分码元为 $e_1,$

e_2, \dots ; 如果游长编码表中第1部分码元 $S_e = 0, S_o = 0$, 则游长编码表中第2部分与第3部分码元均为空,第4部分为原始游长子序列 L_j 所对应的二值码子序列。

依次拼接游长子序列 L_j 多次缩减后游长编码表,生成游长序列 L 多次缩减后游长编码表,记为 T 。

表1 游长子序列 L_j 多次缩减后游长编码表

$L_j(s t)$ 对应的二值码子序列									
S_e	S_o	s	t	e_1	e_2	\dots	o_1	o_2	\dots

(4) 游长序列 L 的转换关系码表 R 。

依次将游长子序列 L_j 经倒序置换后的转换关系码表 R_j 置入 R 中,生成游长序列 L 的转换关系码表 R 。

(5) 二元序列压缩编码表。

拼接码表 T 和 R ,生成二元序列压缩编码表。

2 二元序列无损压缩数值实验

文中选取了三幅二值图像(a)、(b)、(c)及一幅256级灰度图像(d)作为测试对象(见图1)。其中,图像(d)按其灰度深度8,依位平面分解为8个二值图像。文中采用逐行抽样的方式对四幅图像进行处理,分别生成四组二源码序列。表2给出了文中方法与CCITT.4(G3)方法的压缩比。

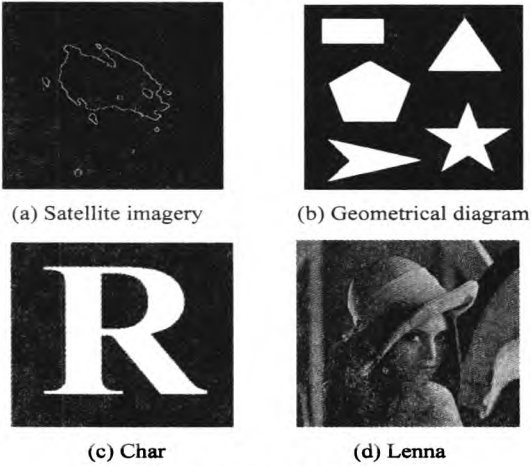


图1 测试图像

表2 实验结果比较

	Image (a)	Image (b)	Image (c)	Image (d)
Size	256×256	256×256	256×256	256×256
CCITT.4(G3)	5.95	4.00	3.73	0.74
Proposed	15.86	11.46	12.94	1.24

实验结果表明,文中给出的游长多次缩减方法对于二元序列的无损压缩相比CCITT.4(G3)具有较高的压缩比。图(d)的8个位平面中仅有较高的3个位平面所对应的二元序列游长较长,可实现编码压缩,其余较低的5个位平面所对应的二元序列游长较短,按CCITT.4(G3)方法对这5个位平面所对应的二元序列进行编码,导致数据严重膨胀,淹没掉整体的数据压缩量,致使压缩比低于1,未实现数据压缩。而采用本方

法后,对较低的 5 个位平面不重新编码,仅增加每个游长子序列的 2 位压缩状态标记 S_e 和 S_o ,从而实现对较高 3 个位平面压缩的同时,避免了较低的 5 个位平面重新编码后导致的数据膨胀,压缩比大于 1,实现了数据压缩。

游长多次缩减编码方法对二元序列的编码主要依据游长缩减值和游长转换关系码元,所以解码时仅需要通过游长缩减起始值和相应的转换关系码元,通过游长逐次复原,即可恢复原始游长序列,而无需类似 CCITT.4(G3)方法,按游长编码表逐一匹配进行解码,故此,文中提出的二元序列游长多次缩减无损压缩编码方法也是一种运行效率较高的编码方法。

3 结束语

文中提出一种具有自适应性的二元序列游长多次缩减的无损压缩编码方法,采用试探法选取游长子序列中奇偶游长缩减起始值,基于游长缩减起始值对较长游长进行缩减,而对小于游长缩减起始值的游长不进行游长缩减处理(即对较短游长保留其原始游长,不重新进行游长编码),解决了较短游长重新编码后导致的数据膨胀。此外,文中提出的方法可对游长子序列进行多次游长缩减处理,直至达到最佳压缩效果。通过与国际传真标准 CCITT.4(G3)进行试验对比,证明了文中提出的方法对二元序列的压缩具有更高的压缩比。

参考文献:

- [1] 沈世铨,陈鲁生. 信息论与编码理论[M]. 北京:科学出版社,2002:11-27.
- [2] 张海燕,王东木,宋克欧,等. 图像压缩技术[J]. 系统仿真学报,2002,14(7):831-835.
- [3] 张春田,苏育挺,张 静. 数字图像压缩编码[M]. 北京:清华大学出版社,2006.
- [4] Berghorn W, Boskamp T, Lang M, et al. Fast Variable Run-length Coding for Embedded Progressive Wavelet-based Image Compression[J]. IEEE Transactions on Image Processing, 2001,10(12):1781-1790.
- [5] 韩玉坤. 数字图像压缩方法研究[J]. 计算机与信息技术, 2007(16):324-325.
- [6] Papadonikolakis M E, Kakarountas A P, Goutis C E. Efficient high-performance implementation of JPEG-LS encoder[J]. Journal of Real-time Image Processing, 2008(3):303-310.
- [7] Feng Xi. Comparison research of some image lossless compression algorithm[D]. Beijing: Graduate School of Chinese Academy of Sciences, 2008.
- [8] Stabno M, Wrembel R. RLH: Bitmap compression technique based on run-length and Huffman encoding[J]. Information Systems, 2009,34(4-5):400-414.
- [9] 祝本明,刘桂华. 一种改进的游程编码算法[J]. 西南科技大学学报,2007,22(3):75-78.
- [10] 崔春晖. 基于二次游程编码和动态码表截取的变长编码方法[D]. 武汉:华中科技大学,2006.
- [11] 孟 君. 文档图像压缩的 RLC 实现[J]. 贵州工业大学学报(自然科学版),2002,31(4):105-106.

(上接第 30 页)

- 1) 角点提取的精度有误差;
- 2) 目标物体不是标准的长方体,会有一定误差。

5 结束语

文中是结合了传统的张氏平面模板标定法,在家庭服务机器人平台上,用 OPENCV 和 VC++ 6.0 完成了双目摄像头的标定与目标物体的定位。实验结果表明该标定方法有效,定位结果可以满足要求,双目摄像机的成功标定和目标物体定位为后续的机器人控制做了很好的铺垫。

参考文献:

- [1] 张宗华,彭 翔. 实物表面三维相位图定标的研究[J]. 仪器仪表学报,2000,21(5):539-541.
- [2] 贺 振,李琼砚. 基于立体视觉的摄像机标定方法的研究[J]. 科技创新导报,2012(15):16-17.
- [3] Zhang Z Y. A flexible new technique for camera calibration[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000(11):1330-1334.
- [4] 吕朝辉,张兆杨. 基于神经网络的立体视觉摄像机标定[J]. 机械工程学报,2003(9):93-95.
- [5] Ma S. A self-calibration technique for active vision system[J]. IEEE Trans. on Robotics and Automation, 1996(12):114-118.
- [6] 许俊勇,王景川,陈卫东. 基于全景视觉的移动机器人同步定位与地图创建研究[J]. 机器人,2008,30(4):289-297.
- [7] 张伟华,陈 军. CCD 摄像机标定[J]. 传感器与微系统, 2009,28(20):107-109.
- [8] 林慧英,苏 建,刘玉梅,等. 基于图像畸变矫正的摄像机标定方法[J]. 吉林大学学报,2007,37(2):433-437.
- [9] Weng J, Cohen P, Herniou M. Camera calibration with distortion models and accuracy evaluation[J]. IEEE Trans. on PAMI, 1992,14(10):965-980.
- [10] 王科俊,魏 娟. 基于双面圆的双目视觉分布定位法[J]. 应用科技,2010,37(1):36-40.
- [11] 尚 倩,阮秋琦. 双目立体视觉的目标识别和定位[J]. 智能系统学报,2011,8(4):303-308.
- [12] 周子啸,赵晓琳. 基于仿真机器人的双目视觉定位[J]. 电视技术,2010(8):106-109.

二元序列游长多次缩减的无损压缩编码方法

作者: 高健, 饶珺, 孙瑞鹏, GAO Jian, RAO Jun, SUN Rui-peng
作者单位: 上海大学机电工程与自动化学院, 上海, 200072
刊名: 计算机技术与发展 
英文刊名: Computer Technology and Development
年, 卷(期): 2013, 23 (6)

本文链接: http://d.g.wanfangdata.com.cn/Periodical_wjz201306008.aspx