

海洋湍流数据实时压缩方法研究

周丽芹, 葛安亮, 王向东, 李坤乾, 宋大雷

(中国海洋大学 工程学院, 山东 青岛 266100)

摘要: 海洋湍流因具有随机性特点, 目前多采用统计学理论进行研究, 因此需要获取大量的湍流观测数据, 这给湍流观测设备的数据存储和传输带来挑战。针对上述问题, 本文在分析海洋湍流数据特征的基础上, 提出了一种高效实时的无损数据压缩方法。以大量的湍流数据增量信息作为数据源构建霍夫曼编码表, 并以此作为湍流压缩和解压的字典, 从而提高了压缩效率。通过对历史海洋湍流数据进行压缩实验, 证明该方法的湍流数据压缩比低至 25%, 并且具有压缩速度快、处理器占用率低等特点。

关键词: 海洋湍流; 数据压缩; 霍夫曼编码

中图分类号: P715.9; TP301.6 文献标识码: A 文章编号: 1000-3096(2019)02-0027-07

DOI: 10.11759/hyqx20180801001

湍流在海洋能量和水体的交换演化中起着非常重要的作用^[1]。研究发现, 海洋湍流的脉动信号频率通常介于 1~100 Hz^[2]间, 为了更真实还原海洋湍流数据, 湍流采集设备的采样频率通常高于 1 000 Hz, 如 sea&sun 公司的 MSS(Mikro-Struktur-Sonde)系列剖面仪, 其湍流采样频率为 1 024 Hz^[3], Rockland Scientific 公司的 VMP(Vertical Microstructure Profiler) 6000 湍流剖面仪, 最高采样频率可达 4 096 Hz^[4]。湍流设备高频采样会带来数据量庞大的问题, 如面向全球深海大洋的智能浮标项目, 设备在位工作 3 年间, 湍流传感器将产生 100 GB 以上的观测数据^[5]。同时海洋观测设备为提升续航能力一般采用低功耗微控制器, 数据处理能力较弱, 而且大量的观测数据也会给设备的数据存储与传输带来巨大压力, 既增加了海洋观测设备的成本, 也限制了其续航能力。

高效的数据压缩方法可以有效地解决这一问题, 但对于数据量庞大的海洋湍流数据, 目前还没有实时压缩方法。在其他领域, 数据压缩方法已经得到广泛地应用, 相关研究主要集中在行程编码、霍夫曼编码、算术编码和 LZW(Lempel-Ziv-Welch Encoding) 等数据压缩方法上。基于霍夫曼算法以及 LZW 算法, 胡斌等^[6]与王文彦等^[7]分别研究了面向海洋观测浮标数据压缩方法, 有效降低了浮标数据通信成本。付东洋等采用优化的行程编码压缩算法对卫星海洋遥感数据进行研究, 极大地节约了海洋遥感数据的存储与共享发布空间^[6]。李浩琦等^[9]在 CUDA(Compute Unified Device Architecture)架构上利用 GPU (Graphics

Processing Unit)平台对海洋监测数据采用霍夫曼并行压缩方式, 实现了海底观测网中监测数据的实时压缩。此外, 单海艳等^[10]提出了面向 CTD (Conductance, Temperature, Depth)海洋观测数据的压缩加密方法。赵忠生等^[11]将数据压缩算法用于深海潜标 ADCP(Acoustic Doppler Current Profilers)的数据传输任务中。Lee 等^[12]对周期性的心电信号进行预处理后利用霍夫曼编码, 实现了心电数据的实时压缩和传输。朱凌云等^[13]采用一阶差分结合霍夫曼和 LZ77 (由 Abraham Lempel 和 Jacob Ziv 于 1977 年提出)压缩算法, 在嵌入式平台上实现了心电数据的高效压缩。Medeiros 等^[14]利用统计数据构建霍夫曼字典对无线传感器网络中的温度和湿度数据实现了轻量级数据压缩。

总的来说, 行程编码对于长重复值字符串的数据压缩效果良好, 但不适用于随机性强的海洋湍流数据的压缩; LZW 压缩、霍夫曼编码对系统资源占用较大, 算术编码依据统计原理对源数据编码, 对处理器

收稿日期: 2018-07-31; 修回日期: 2018-11-08

基金项目: 国家自然科学基金重大科研仪器研制项目(41527901); 中央高校基本科研业务费专项(201813022)

[Foundation: National Major Scientific Instruments Development Program of China, No. 41527901; Fundamental Research Funds for the Central Universities, No. 201813022]

作者简介: 周丽芹(1972-), 女, 山东青岛人, 副教授, 主要研究领域为嵌入式技术, 海洋智能观测, 电话: 0532-66781109, E-mail: liqin72@126.com; 李坤乾, 通信作者, 男, 讲师, 主要研究领域为计算机视觉、水下机器人信号处理及控制, 电话: 0532-66781786, E-mail: likunqian@ouc.edu.cn

计算精度有较高要求,它们均不利于海洋观测设备的微型化和低功耗化。因此,研究针对海洋湍流数据的高效实时无损压缩方法,不仅可以提高湍流数据压缩率,降低湍流数据的存储与传输成本,对于提升海洋观测平台的长续航观测能力也具有重要意义。

本文分析了海洋湍流的数据特点,对大量的湍流数据的增量信息进行了统计,统计结果表明,湍流数据的增量信息非常集中,可应用霍夫曼编码进行数据压缩。为提高压缩效率,设计了基于增量的字典霍夫曼编码方法,压缩实验结果表明,该方法具有数据压缩比低,压缩时间短,资源占用率低的特点,可广泛应用于实时性要求高、存储资源有限的海洋观测设备中。

1 湍流数据介绍

1.1 湍流数据获取

本文中使用的海洋湍流数据为2017年3月采用实验室自研的自容式湍流剖面观测仪^[15]于南海采集的试验数据,自容式湍流剖面观测仪共搭载2个垂直向分布的德国PNS(Prandke、Neumann、Stips)系列剪切流传感器,剪切流传感器获得的湍流模拟信号依次通过高通放大和低通滤波后经A/D转换获得数字信号,获取方法如图1所示:

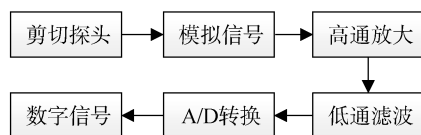


图1 湍流数据的获取流程
Fig. 1 Flowchart of turbulence data acquisition

自容式湍流剖面仪采用的16位A/D模数转换芯片,采样范围 ± 2.5 V,采样频率为1 024 Hz,每一帧数据5个字节,每通道数据占用2个字节(高字节、低字节),其中帧头为0x37,数据格式如表1所示。每隔3 min创建一个TXT文件用于存储湍流数据,因此每个TXT文件的大小约为900 Kb。

表1 湍流数据格式
Tab. 1 Format of turbulence data

帧头	剪切 1		剪切 2	
0x37	高字节	低字节	高字节	低字节

截取自容式湍流剖面观测仪海试试验中某采样周期的3 min湍流数据,将其数字信号还原为电压信

号后的数据曲线如图2所示,电压信号在一定范围内波动。

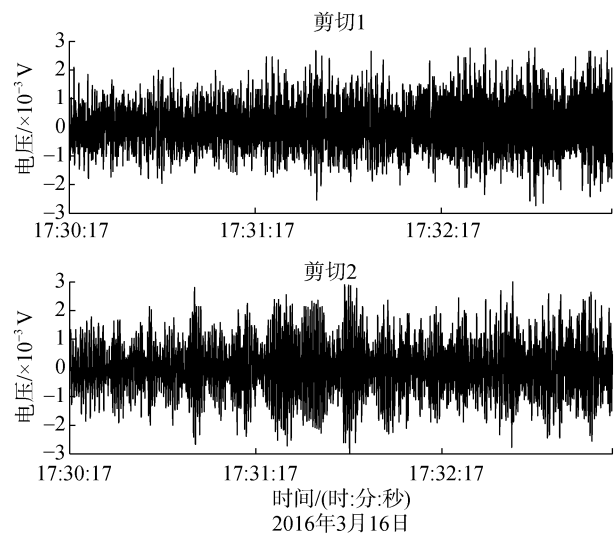


图2 2016年3月16日湍流信号电压曲线
Fig. 2 Voltage curve of turbulence signal

1.2 湍流数据特征分析

以图2中的二维剪切数据为例进行统计分析,得到剪切数据特征如表2所示。可以看出,海洋湍流数据较为集中、波动较小。

表2 剪切数据特征
Tab. 2 Shear data features

	平均峰峰值/ 10^{-3} V	方差/ 10^{-8} V ²	标准差/ 10^{-4} V
剪切 1	5.5	6.50	8.06
剪切 2	6.0	8.52	9.23

对7 000多万连续海洋湍流数据 S_n 进行差值处理 $\Delta S_n = S_n - S_{n-1}$,获取相邻采样点的数据增量信息,并统计海洋湍流数据增量 ΔS_n 的出现次数,结果如图3所示。

通过计算得出:

- (1) 99%以上湍流数据增量集中在 $-44 \sim 44$ LSB (least significant bit)之间;
- (2) 10%以上的湍流数据增量为零。

因此,海洋湍流数据的增量集中,可应用霍夫曼编码进行数据压缩。

2 基于增量的字典霍夫曼编码

2.1 霍夫曼编码介绍

霍夫曼(Huffman)编码,是霍夫曼于1952年提出

的一种可变字长的编码算法^[17]。霍夫曼编码使用自下而上的方式，根据信息中字符出现的频率构建二叉树，出现频率高的字符码字短，出现频率低的字符码字长，用最少的码长重新编码源数据，因此也称为最优编码。通过对信源进行霍夫曼编码，使编码后的数据的平均长度变小，期望值降低，从而达到数据的无损压缩效果^[16]。

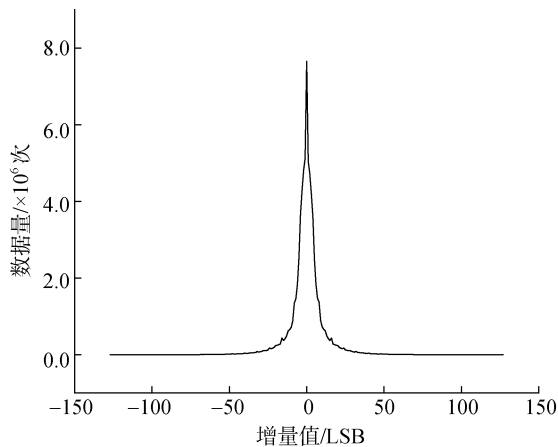


图 3 湍流增量统计
Fig. 3 Turbulence increment statistics

霍夫曼编码的核心是霍夫曼树的构建，霍夫曼树的构建及编码的方法如下。

首先统计信源中每个字符的出现次数，并将字符出现的次数按照从小到大的顺序依次排列。

然后把出现次数最少的两字符的次数值相加得到新节点，将这两个次数最少的字符作为新节点的子节点。其中，这两个子节点中次数较少的节点作为左节点，次数较多的节点作为右节点。将此新节点按顺序排列到队列中，生成新的队列。重复此操作生成二叉树，直到最后相加得到的次数值为总次数值。

最后将二叉树中每个父节点的左节点指定为 0，右节点指定为 1。

至此，已自下而上地构造了一棵霍夫曼树，再按照自上而下的顺序得到每个字符的编码，最终获得编码表。

由于霍夫曼编码是基于概率统计结果对信源进行编码，必须依次读取两遍信源数据：第一遍先生成霍夫曼二叉树得到编码表，第二遍按照编码表对信源进行编码。这种数据编码方法较为高效，但是也存在弊端：对数据多次读取和编码，会占用大量的处理器资源，这无疑提高了处理器性能需求和处理功耗，不利于数据处理设备的微型化和低功耗化。

2.2 增量编码实现

根据海洋湍流数据波动范围小的特点，且通过大量的统计发现相邻两个采样点的数据增量大多在 $-127 \sim 127$ LSB 之间，增量可用 1 个字节表示，因此可以通过计算两个相邻采样点的数据差值，并去除每帧数据的帧头，只保存湍流数据的增量信息，从而降低数据的熵。

部分湍流数据也存在幅值变化剧烈的现象，增量值可能在 $-127 \sim 127$ LSB 之外，对于这种情况，在数据存储过程中添加一个字节标志位 0x80，并保存源数据，表示不对数据进行差值处理。为避免标志位与数据存在冲突，选用有符号 byte 值 0x80 作为标志位，其十进制数值为 -128 ，所以在处理后所得的数据中，所有的 0x80 只表示为其后两字节数据为没有进行差值运算的源数据。数据增量编码压缩处理流程如图 4 所示。

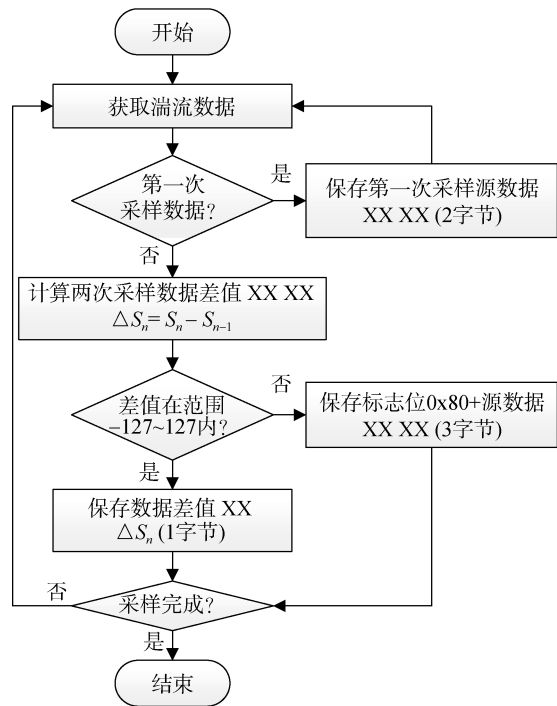


图 4 数据增量编码压缩处理流程
Fig. 4 Flowchart of incremental data coding compression

使用数据增量编码压缩方法对历史湍流数据进行实时数据压缩，当差值在 $-127 \sim 127$ LSB 之间时，每一帧 5 个字节的湍流数据可压缩为 2 个字节，因此理论上的极限压缩比为 40%。而且在数据编码压缩过程中，只对原始数据进行提取和差值计算，数据处理方法简单高效，并可将压缩后的数据完全真实

还原,因此增量编码属于数据无损压缩方法。

2.3 基于增量的字典霍夫曼编码实现

根据研究和试验表明,2.1 节中的霍夫曼编码方法对处理器性能要求很高,在低功耗的嵌入式湍流观测设备上难以实现,所以还需要对当前算法进行改进和优化;2.2 节中的增量编码方法压缩效率还有提高的空间。

由于霍夫曼编码最核心的部分是霍夫曼编码表的获取,也是造成霍夫曼编码过程中计算量过大的主要原因。在对大量的湍流数据增量信息统计时发现,湍流数据增量值越小,其出现的概率越高(图 3)。基于上述分析,本节提出以湍流数据增量信息作为数据源,通过霍夫曼压缩算法预先构建霍夫曼树,并以此作为湍流压缩和解压的字典。在湍流数据压缩过程中不再动态地构建霍夫曼树,极大地缩短了湍流数据的压缩时间,增强了系统的实时性。霍夫曼字典的构建、数据压缩和解压过程如图 5 所示。

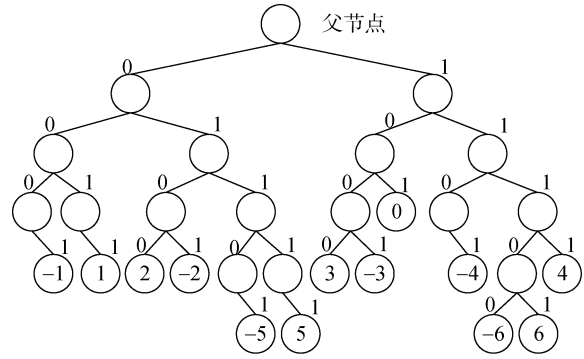


图 6 部分霍夫曼树示意图

Fig. 6 Schematic of parts of Huffman tree

表 3 部分霍夫曼编码字典

Tab.3 Excerpts from Huffman coding dictionary

差值(HEX)	差值(DEC)	编码	码长
00	0	101	3
01	1	0011	4
02	2	0100	4
03	3	1000	4
04	4	1111	4
05	5	01111	5
06	6	11101	5
...
FA	-6	11100	5
FB	-5	01101	5
FC	-4	1101	4
FD	-3	1001	4
FE	-2	0101	4
FF	-1	0001	4

注: HEX 表示十六进制计数制; DEC 表示十进制计数制。

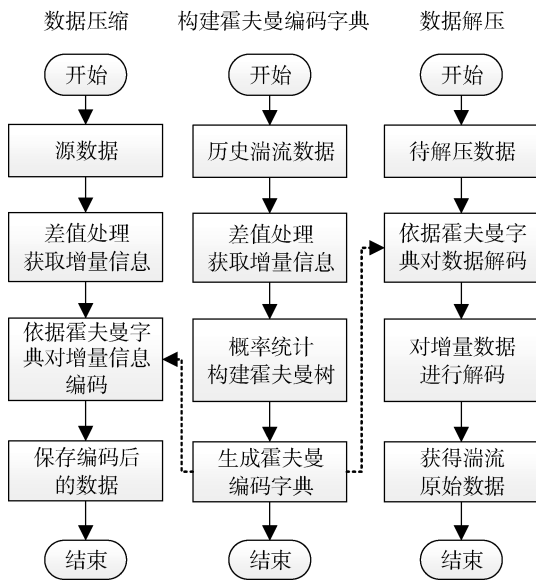


图 5 霍夫曼字典的构建、数据压缩和解压过程

Fig. 5 Flowchart of Huffman dictionary construction, data compression and decompression

该方法构建的部分霍夫曼树示意图如图 6 所示,由于篇幅有限,所展示的霍夫曼二叉树只包含-6~6 节点。

该方法生成的霍夫曼编码字典如表 3 所示,数据差值范围为 0x00~0xFF,共 256 个,编码的码长一般在 3~18 位之间。因此每个差值的编码最多占用 3 个字节,整个编码字典的大小不超过 1 KB。

其中,霍夫曼编码字典的构建是在性能较高的 PC 电脑端实现,生成的霍夫曼编码字典占用 1 KB 空间,将该字典导入低功耗嵌入式单片机系统,配合传感器实现对海洋湍流数据的实时压缩。

3 结果与分析

3.1 不同压缩方式结果对比

使用第 2 节中的 3 种方法和当前流行的压缩软件对 200 组(约 174 MB)海洋湍流历史数据进行压缩,得到 200 组海洋湍流数据的平均压缩比结果如表 4 所示。其中,本文方法以所有被压缩数据的增量信息为信源生成霍夫曼树。

表 4 压缩结果对比

Tab. 4 Compression ratios of different compression methods

压缩方式	压缩比/ %	压缩 时间/s	解压 时间/s
7Z	30.40	56.82	5.71
ZIP	44.98	51.64	3.23
RAR	32.94	11.47	5.21
增量编码	40.05	8.463	10.27
增量编码+7Z	24.34	29.353	14.29
增量编码+ZIP	25.24	41.963	12.64
增量编码+RAR	26.06	22.583	15.2
增量编码+霍夫曼	24.37	24.269	31.72
增量编码+字典霍夫曼	25.47	23.054	31.63

注：由于使用的压缩软件不同，为了使压缩时间具有一定的参考性，7Z 和 ZIP 压缩方式为 7-Zip 软件在单一线程下使用软件默认参数进行压缩，7Z 字典大小 64 MB，ZIP 字典大小 32 KB；RAR 压缩方式为 WinRAR 软件以默认参数进行压缩（软件无法设置单一线程下压缩），字典大小 32 MB；增量编码、霍夫曼和字典霍夫曼均为 C# 平台编写的压缩程序在单一线程下压缩。软件运行平台均为 64 位 WIN10 系统，I7-4700HQ 处理器。

其中，压缩比是评判系统数据压缩性能的重要参数之一，指数据被压缩的比例，用字母 R 表示，通常 $R < 1$ ，压缩比 R 越小，表示压缩效果越好。计算公式为：

$$R = \frac{\text{压缩数据长度(比特数)}}{\text{原始数据长度(比特数)}} \quad (1)$$

通过表 4 中的 7Z、ZIP、RAR 和增量编码压缩方式可以看出，对海洋湍流数据采用直接压缩处理方式，压缩效果并不理想，压缩比较高。对海洋湍流数据在增量编码的基础上再次压缩，可将压缩比降至 25% 左右，压缩效果非常明显。基于增量编码的 RAR 压缩方法虽然速度较快，但其压缩方法并未开源；基于增量编码的 7Z 压缩方法和基于增量编码的霍夫曼压缩方法的压缩比最低，对系统存储资源占用最小，更适用于数据量庞大的海洋湍流数据压缩，但是这两种方法对处理器的运算性能要求较高，对 RAM(内存)需求较大，而低功耗、低性能的嵌入式系统的内存空间通常只有几十 KB(如 STM32F103 系列单片机内存空间在 6~96 KB 之间)；基于增量编码的字典霍夫曼压缩方法的压缩比非常接近基于增量编码

表 6 压缩结果

Tab. 6 Compression results

压缩方法	平均压缩时间/ms	CPU 占用率/%	内存占用/KB	压缩比/%
增量+字典霍夫曼	1674	0.93	1	25.66
增量+霍夫曼	3780	2.10	361	24.59

注：表 6 中该两种压缩方法在单片机端实现，表 4 中的压缩方法在 PC 端实现，两种平台的压缩算法实现略有差异，因此造成该两种压缩方法在两种平台上的压缩时间差不同。本项实验中，本文方法以所有 20 组被压缩数据的增量信息为信源生成霍夫曼树。

的霍夫曼压缩方法，且对 CPU 和内存占用低，可以容易的移植到低功耗、低性能的海洋仪器中，从而提高了仪器的数据存储量，增强了海洋仪器的续航能力。

3.2 不同霍夫曼树压缩结果对比

为了分析基于增量的字典霍夫曼压缩方法中不同霍夫曼树对压缩结果的影响，将 200 组数据按时间顺序均分为 A、B 两部分，使用 B 组的 100 组数据构建霍夫曼编码表对 B 组数据进行压缩的压缩比为 25.80%，可认为该压缩比为 B 组数据基于增量的字典霍夫曼压缩方法的最低压缩比。

从 A 组中按固定间隔分别多次取 1 组、2 组、4 组、5 组、10 组、20 组、50 组、100 组数据构建霍夫曼编码字典，并以此对 B 组数据进行编码压缩，得到的数据压缩比结果如表 5 所示。

表 5 数据压缩比结果

Tab. 5 Data compression ratios

组数	最小/%	最大/%	平均/%
1	27.15	32.24	30.49
2	25.87	30.09	27.09
4	25.89	27.55	26.52
5	25.82	30.07	26.94
10	25.81	25.84	25.83
20	25.81	25.83	25.82
50	25.81	25.81	25.81
100	25.81	25.81	25.81

由表 5 可以看出，选用不同的历史数据所得到的数据压缩比差距较小，由于湍流数据增量较为集中(1.2 节)，即使采用较少的历史数据，数据压缩比也能达到 30% 左右，且采用的历史湍流增量数据越多，平均压缩比越接近最低压缩比。

3.3 单片机压缩实验

将基于增量的字典霍夫曼压缩方法和基于增量的霍夫曼压缩方法移植到 STM32L476(80 MHz 主频，1MB Flash，128KB SRAM)单片机上，并对 20 组海洋湍流数据(180 s 一组，每组数据约 900 KB)进行压缩，压缩实验结果如表 6 所示。

由表 6 可以看出, 基于增量的字典霍夫曼编码方法更为高效, 表现为较低的数据压缩比和更低的处理器资源占用率。由 2.1 节, 霍夫曼编码需要读取两遍源数据, 因此 900 KB 的湍流数据需要 360 KB 的内存空间用于缓存其增量数据, 这对于内存资源极为有限的低功耗单片机极为致命。而基于增量的字典霍夫曼编码方法只读取一遍源数据, 且仅需 1 KB 内存空间用于存储编码字典, 因此更为高效。

综上, 基于增量的字典霍夫曼编码方法不仅可以节省数据存储空间, 而且可以大大降低对系统处理器性能需求。更重要的是, 该方法为无损编码方法, 可通过对压缩后的数据进行无损解压得到源数据。因此, 基于增量的字典霍夫曼编码方法可对数据量庞大的海洋湍流数据实现高效实时压缩, 该方法的优点也决定了其适用于低功耗、低性能的嵌入式海洋仪器, 也可推广至其他具有数据量庞大、数据波动幅值小、增量信息集中等特点的数据处理系统中。

4 结论

本文通过对自容式湍流剖面观测仪获取的海洋湍流数据进行分析, 获得了海洋湍流数据增量统计学特征, 针对数据量庞大的海洋湍流数据在海洋设备中存储与传输难题, 提出了基于增量的字典霍夫曼编码方法。实验表明, 该方法可在嵌入式单片机系统上实现海洋湍流数据的高效实时无损压缩, 且压缩比低至 25%。因此基于数据增量统计的霍夫曼压缩方法可广泛应用于需要存储或传输海洋湍流数据的仪器或系统, 极大地降低了海洋观测设备的数据存储与传输成本。

参考文献:

[1] 康建军, 邬海强, 杨庆轩, 等. 海洋湍流观测技术[J]. 海洋技术学报, 2007, 26(3): 19-23.
Kang Jianjun, Wu Haiqiang, Yang Qingxuan, et al. Ocean Turbulence Measurement Technology[J]. Ocean Technology, 2007, 26(3): 19-23.

[2] 田川. 深海湍流混合长期连续观测技术与应用研究[D]. 天津: 天津大学, 2012.
Tian Chuan. Study on long-term turbulence monitoring technique and its application in deep-sea mixing research[D]. Tianjin: Tianjin University, 2012.

[3] Sea & Sun Marine Tech. Microstructure Probes: the MSS-family[EB/OL]. <http://www.sea-sun-tech.com/marine-tech/oceanography/mss-microstructure-probe.html>, 2018-7-30.

[4] Rockland Scientific. VMP 6000 Deep-Sea Vertical Microstructure Profiler[EB/OL]. <https://rocklandscientific.com/products/profilers/vmp-6000/>, 2018-7-30.

[5] 贾鹏飞. 智能浮标物理接驳盒的设计与开发[D]. 青岛: 中国海洋大学, 2018.
Jia Pengfei. Design and Development of Physical Junction Box Based On Smart Float[D]. Qingdao: Ocean university of China, 2018.

[6] 胡斌, 李忠强, 刘婷婷, 等. Huffman 与 LZW 算法在海洋观测浮标通信数据压缩中的应用研究[J]. 海洋科学, 2018, 42(1): 6-10.
Hu Bin, Li Zhongqiang, Liu Tingting, et al. Application of Huffman and LZW algorithms in data compression for ocean-observation-buoy communication[J]. Marine Sciences, 2018, 42(1): 6-10.

[7] 王文彦, 李文庆, 王晓燕, 等. 数据压缩技术在海洋资料浮标通信系统中的应用[J]. 山东科学, 2015, 28(2): 1-5.
Wang Wenyan, Li Wenqing, Wang Xiaoyan, et al. Application of data compression in the communications system of an oceandata buoy[J]. Shandong Science, 2015, 28(2): 1-5.

[8] 付东洋, 李宇隆, 侯骏雄, 等. 卫星海洋遥感数据的无损压缩及解压算法[J]. 海洋技术学报, 2012, 31(2): 45-48.
Fu Dongyang, Li Yulong, Hou Junxiong, et al. Lossless compression and decompression algorithms for satellite ocean remote sensing data[J]. Ocean Technology, 2012, 31(2): 45-48.

[9] 李浩琦, 王海斌, 汪俊. 基于 CUDA 的海洋监测数据并行压缩技术研究[J]. 海洋技术学报, 2014, 33(1): 39-44.
Li Haoqi, Wang Haibin, Wang Jun. Study on the parallel compression technique for marine monitoring data based on CUDA[J]. Ocean Technology, 2014, 33(1): 39-44.

[10] 单海艳, 熊学军, 郭延良, 等. CTD 海洋观测数据多重压缩加密算法设计[J]. 海岸工程, 2018, 37(2): 23-29.
Shan Haiyan, Xiong Xuejun, Guo Yanliang, et al. Design of multiple compression encryption algorithm for CTD ocean observation data[J]. Coastal Engineering, 2018, 37(2): 23-29.

[11] 赵忠生, 袁志伟, 黄磊, 等. 深海潜标 ADCP 的实时数据传输[J]. 海洋科学, 2012, 36(8): 94-97.
Zhao Zhongsheng, Yuan Zhiwei, Huang Lei, et al. Real-time data transmission of the ADCP installed on subsurface mooring system[J]. Marine Sciences, 2012, 36(8): 94-97.

[12] Lee S, Kim J, Lee M. A real-time ECG data compression and transmission algorithm for an e-health device[J]. IEEE Transactions on Bio-medical Engineering, 2011, 58(9): 2448.

- [13] 朱凌云, 李世星, 林红胜. 一种基于 S3C2440 的嵌入式实时心电图数据压缩算法[J]. 北京生物医学工程, 2012, 31(3): 258-261.
Zhu Lingyun, Li Shixing, Lin Hongsheng. A real-time ECG data compression algorithm based on S3C2440[J]. Beijing Biomedical Engineering, 2012, 31(3): 258-261.
- [14] Medeiros H P, Maciel M C, Demo Souza R, et al. Lightweight Data Compression in Wireless Sensor Networks Using Huffman Coding[J]. International Journal of Distributed Sensor Networks, 2014, 2014(1): 1-11.
- [15] 乜云利. 自容式湍流剖面观测仪设计与试验研究[D]. 青岛: 中国海洋大学, 2015.
Nie Yunli. The design of self-contained profiler turbulence observation instrument and test research[D]. Qingdao: Ocean University of China, 2015.
- [16] 夏萍. 数据压缩技术的研究[D]. 太原: 中北大学, 2010.
Xia Ping. Research on data compression technology[D]. Taiyuan: North University of China, 2010.
- [17] Huffman D A. A method for the construction of minimum-redundancy codes[J]. Proceedings of the IRE, 1952, 40(9): 1098-1101.

Research on realtime compression of ocean turbulence data

ZHOU Li-qin, GE An-liang, WANG Xiang-dong, LI Kun-qian, SONG Da-lei
(Ocean University of China, College of engineering, Qingdao 266100, China)

Received: Aug. 01, 2018

Key words: ocean turbulence; data compression; Huffman code

Abstract: Because of the randomness of ocean turbulence, most related studies of it are based on statistical theory. It is necessary to obtain a large amount of turbulence observation data, which brings us enormous challenge to data storage and transmission for turbulence observation equipment. In this study, an efficient real-time lossless data compression method, based on the analysis of the characteristics of ocean turbulence data, is proposed properly. The Huffman coding table constructed with a large amount of turbulence data increment information is used as a dictionary for turbulence compression and decompression. The above characteristic makes this method possess a high compression efficiency. Our experimental results indicate that the compression ratio of turbulent data can be as low as 25% through compression experiment of historical ocean turbulence data using this efficient method, and this method also has the characteristics of fast compression speed and low processor occupancy rate.

(本文编辑: 李晓燕)