Digital Signal Processing 69 (2017) 204–211

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

Lossless coding scheme for data acquisition under limited communication bandwidth

Yazhou Yuan [a], Yu Zhang [a], Zhixin Liu [a,*], Xinping Guan [b]

[a] Institute of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China
[b] School of Electronic, Information and Electrical Engineering, Shanghai Jiaotong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Available online 12 July 2017

Keywords:
Data compression
WDAS
Lossless Coding
Limited bandwidth

ABSTRACT

Wireless data acquisition system (WDAS) is widely used in many application fields. Due to the conflict between bandwidth constraint and large amount of real-time data with high sampling frequency or multi-channel sampling, it is a challenging problem to transmit data to the back-end processing server timely and effectively. A novel and simple lossless source coding scheme, called PVA (Pre-processing and Valid word length Adaptive coding), is proposed for WDAS under limited communication bandwidth. The proposed coding scheme improves the utilization of communication bandwidth and realizes the real-time compression and transmission by the technique of common binary fields finding, repeating fields reducing and data stream recombination. PVA scheme also improves the precision of the measured signals compared to the normal oversampling method under the same bandwidth. It is thus effective for the applications which are in high demand of fidelity. The experiments on vibration and electroencephalography (EEG) signals show that the proposed PVA scheme is effective and valuable in WDAS.

1. Introduction

Data acquisition system (DAS) has been widely used in many fields such as bridge health diagnosis, power quality analysis, body health monitoring [1]. Since using DASs with long cable to transmit data has high cost and low flexibility, more and more DASs are turning to use wireless technology [2,3]. DASs collect corresponding signals with high sample frequency and multi-channel, such as mechanical equipment diagnosis fields including bearings, gears, medical signals including electroencephalography (EEG) and electrocardio (ECG) [4,5]. Especially, in modern EEG, the number of channels are up to 256, and each channel can sample at 1 Mhz with the resolution up to 32 bits. DASs usually generate a large amount of data [6,7], and if wireless links is adopted, data compression is necessary for the transmission under limited bandwidth of wireless network [8].

Generally, the data compression algorithm is divided into two categories: lossless and lossy algorithm [9]. In this paper, a lossless compression scheme (PVA) based on source coding [10] is proposed to deal with the real time transmission under limited communication bandwidth. The lossless algorithm compresses the original data without losing any information and this is necessary in some specific fields. There are some classic lossless algorithms such as Run Length Encoding (RLE), Huffman, Lempel–Ziv (LZ) coding [1,11–14], and many other correlate improved algorithms [15–19]. However most of these compression methods take effect on the signal in specific fields [20–23]. The PVA scheme proposed in this paper includes two parts: the definition of segment format and the processing of sample data. The segment format uses variable packet length protocol to achieve the best performance of data compression. The processing of sample data is divided into 2 steps: (i) pre-processing based on data correlation, (ii) Valid Word Length (VWL)-Adaptive coding method. The coding scheme can adapt to the different data characteristics and behave better flexibility and adaptiveness, so the compression algorithms provided in this paper is suitable for most of natural signals. To analysis the efficiency and effectiveness of the PVA scheme, the test platform is developed and the real data of vibration and EEG signals are implemented to evaluate the performance of the proposed scheme.

The main features of PVA are summarized as follows.

- Wide range application. There is no special limitation of the original signals and it is especially effective to high-frequency signal. Moreover, PVA is a continuous data stream compression scheme.

* Corresponding author.
E-mail addresses: yzyuan@ysu.edu.cn (Y. Yuan), zhangyu_ysu@hotmail.com, zy-ysu@qq.com (Y. Zhang), lzxauto@ysu.edu.cn (Z. Liu), xpguan@sjtu.edu.cn (X. Guan).

http://dx.doi.org/10.1016/j.dsp.2017.06.028

- Easy implementation. The simple operation and computing are used in the signal processing and it is easy for hardware implementation.
- High compression ratio. PVA shows better performance in aspect of compression ratio and data transmitting efficiency, which is suitable for the scenarios with limited bandwidth.

The rest of this paper is organized as follows: in Section 2, the format of segment and details of PVA compression scheme are introduced; in Section 3, some experiments on PVA scheme are made, and the performance is analyzed. Finally, the conclusions are drawn in Section 4.

2. Basic idea and statement of the problem

2.1. Pre-processing method based on data correlation

2.1.1. Information entropy and redundancy reducing

In this paper, each data collected by the ADC is called an element. For a single sampled element, it is standardized by binary encoding, which is expressed as 0–1 sequence. Assume a 16-bit resolution ADC is used in the system, b is the binary representation of the sampled element and b_m is the m_{th} bit value, p_i stands for the occurrence probability of corresponding b_i as shown in Eq. (1).

$$\begin{bmatrix} b \\ p(b) \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & \dots & b_m \\ p_1 & p_2 & \dots & p_m \end{bmatrix}. \tag{1}$$

Then the information entropy [24] is defined by

$$H_m = -\sum_{i=1}^{m} p_i \ln p_i. \tag{2}$$

It can be seen that more duplicate 0–1 sequence appearing at certain section leads to more redundancy, which means more compressibility. If the occurrence probability of sampling values are equal, a 16-bit binary sequence is necessary to represent a raw element, but considering the correlation among the continuous samples in time domain, the information entropy will be decreased greatly. For a fixed sample frequency, the differences between adjacent sampled data points indicate the gradient of the actual signal, and closer sampling leads to more correlation among the samples. In this paper, a pre-processing method applied on embedded systems is proposed based on the data correlation of the sampled values.

According to the information entropy theory, we can compress the signal by reducing the redundancy. Because natural signals are almost continuous, the adjacent elements have repeating fields. This paper takes the difference operation to reduce the repeating fields. Then we find that most of the elements have common binary sequence in the upper bits as shown in Fig. 1 and Table 1. Bitwise XOR operation is used to reduce the redundancy further.

2.1.2. Difference and bitwise XOR operation

When the data is transmitted by wireless link, a continuous binary sequence of 0 from the Most Significant Bit (MSB) in the processed value could be ignored to save the bandwidth. In order to facilitate the analysis and design, the first binary 1 to the Least Significant Bit is defined as Word Significant Bits (WSB) in this paper, whose length is represented as Valid Word Length (VWL). Raw element sequence read from the Analog to Digital Converter (ADC) are recorded as an array of A, then the difference between every two adjacent raw elements are calculated and stored into array B. The process is shown as Fig. 2.

For natural signals, every two adjacent elements sampled with a proper frequency have approximate value in time domain. By

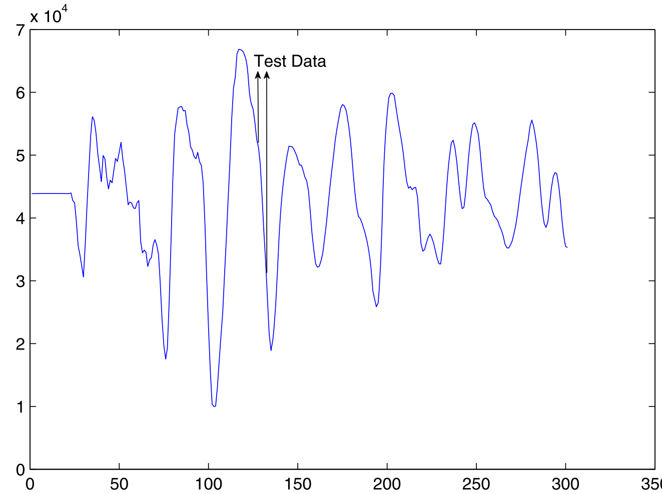

Fig. 1. Vibration data signal.

Table 1
The binary representation of test data.

raw data	difference	after bitwise XOR
0000101110001100	0000101110001100	0000101110001100
0000110011100110	0000000101011010	0000000101011010
0000111001001001	0000000101100011	0000000000111001
0000111110011110	0000000101010101	0000000000110110
0001000011001110	0000000100110000	0000000001100101
0001000111010001	0000000100000011	0000000000110011
0001001010010111	0000000011000110	0000000111000101
0001001100100111	0000000010010000	0000000001010110
0001001110010010	0000000001100001	0000000011110001
0001001111101100	0000000001100100	0000000000000101

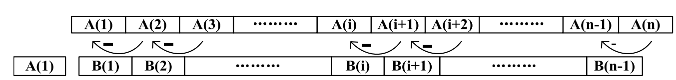

Fig. 2. Schematic diagram of the difference calculated.

calculating the difference, the WSB of $B(i)$ is shorter than $A(i)$ correspondingly. The relationship between the original data array A and difference array B is shown as

$$B(i) = A(i+1) - A(i), i \in (1, n-1). \tag{3}$$

Since the natural signals are not monotonic, after calculating the difference there would be some negative numbers which are usually expressed as two's complement form, and this would make the value of VWL very large. In that case all the negative numbers should be turned to be positive to preserve sufficient compression performance. The process can be expressed as

$$B(i) = B(i) - min(B), \tag{4}$$

where $min(B)$ stands for the min value of array B, and $min(B)$ will be included as a part of the communication frame (refer to the details in Table 2).

The binary representation of each individual element from array B can be expressed as

$$B(i) = sbanr(n-k)\&sbanr(k), \tag{5}$$

where n indicates the valid number bits of each individual element, k represents the different count of bits between adjacent elements. $sbanr(n-k)$ denotes the same binary 0, 1 sequence between the upper bits of $B(i)$ and $B(i+1)$. The symbol '&' represents concatenated binary data for sequence. By using this char-