

An Energy-Efficient Compression Algorithm for Spatial Data in Wireless Sensor Networks

Beihua Ying*

* School of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University, Ningbo, Zhejiang, 315100, China

yingbh@nit.zju.edu.cn

Abstract—Energy efficiency is one of the most important design metrics for wireless sensor networks. As sensor data always have redundancies, compression is introduced for energy savings. In this paper, a lightweight compression algorithm for data with spatial correlation is proposed, which can be implemented on resource constrained nodes to reduce the total energy costs in the whole networks. By adopting pipelining and introducing partial computation, our method achieves a little inter-sensor communication expenditure, and it can reduce energy costs while still keep a good distortion. The simulation results show that, compared with the wavelet compression scheme, the algorithm obtains more energy savings under the same distortion rate.

Keywords—Data Compression, Spatial Correlation, Energy Efficient, Wireless Sensor Networks, Wireless Communication

I. INTRODUCTION

With the advancements in micro-electro mechanical systems and wireless communications, technology developments on wireless sensor networks (WSNs) have been fuelled in recent twenty years. Since the 1990s, when sensor networks emerged as a fundamentally new tool for military monitoring, nowadays they are widely used in many application fields such as agriculture, ecosystems, medical care, traffic control and smart homes, especially for regions which are inaccessible or unattended. By right of the essential function in data collection, WSNs connect the physical environment with human beings [1].

Generally, each sensor node transmits monitoring data over its corresponding path to the sink node. Since they are usually equipped with small batteries and no fixed infrastructure exists, it is prohibitively difficult to replace or recharge them. Energy limitation becomes one of the major differences from other wireless networks. According to this, energy consumption is the primary concern in such networks. Moreover, sensor nodes also suffer from other limited resources, including computation, storage and communication capability. As a result, the major challenge in WSNs is minimizing the energy usage with proper algorithms to prolong the lifetime of the whole networks.

Among many applications, sensor nodes are densely deployed and usually required to collect information for environmental monitoring. In these scenarios, nodes sample surrounding physical phenomena and return the measurements to the sink. The ubiquity of redundancies in raw datasets

inspires researchers to introduce compression technology for reducing data volume and saving communication energy losses. In this paper, an energy-efficient compression algorithm is proposed for spatial redundancy information decorrelation, which is enlightened from a lightweight temporal compression technique named LTC [2]. The method extends data compression to spatial domain, and does well in resolving the problem of information interaction between different nodes, which needs not be considered in the time domain. The test results show that, compared with the wavelet transformation, a common algorithm used in spatial compression, our proposed method can achieve more energy savings under the same distortion rate.

The remainder of this paper is organized as follows. Section 2 presents several related researches on data compression. The implementation process of the proposed algorithm is detailed described in Section 3. Simulation results are presented subsequently with the corresponding conclusions and several future research directions given in Section 5.

II. RELATED WORK

In WSNs, communication is widely considered to be the main power consumptions. More than 90% of energy in sensor nodes are spent on it, such as transmitting, receiving and idle listening [3]. Consequently, in-network data processing has been introduced for energy savings by means of reducing the amount of communication data. Compression is regarded as a traditional technology used in digital communication, broadcasting, storage, and multimedia systems. Being applied to WSNs, data compression faces more new challenges. The biggest one is how to obtain the total energy savings by increasing only little compression energy losses. In other words, a low-cost compression algorithm is urgently needed in WSNs.

Heretofore, most compression algorithms are presented based on time domain. It is relatively simple that sample collection and data compression are performed on the same node. The costs of compression only involve in calculation. Thus, many algorithms have been proposed specifically for WSNs. Most of them are derived from traditional methods, such as predictive compression [4], wavelet transformation [5], data fitting [2] and Huffman coding [6]. Different from the compression in time domain, the raw data for spatial compression is distributed in each node. Due to the information exchange between sensor nodes, the

communication energy costs should be considered, besides the calculation ones. The cooperative work between nodes becomes the key of spatial compression. Several researches are carried out on this point.

One of the data compression in spatial domain is distributed source coding (DSC) [7]. It is a representative technology to minimize the amount of inter-node communication while preserving the data resolution. [8] introduces this coding scheme to a chain-type wireless sensor networks and decreases the decoding error probability with the assistance of convolutional and turbo code. Another research implements DSC in a cluster network [9]. However, the need of pre-know the joint probability density function of the raw data is hard to satisfy, which limits the applicability of DSC [10].

According to the characteristics of data in many WSN applications and the good performances in data stream compression, wavelet transforming is considered as another suitable technology for WSNs. Since the low complexity in processing, a distributed wavelet algorithm, based on the lifting scheme, is proposed in [11] [12]. It introduces partial coefficients to generate 5/3 wavelet coefficients at each node in a one-dimensional network. Likewise, [13] develops a distributed wavelet architecture for a two-dimensional spatial irregular grid topology and extends this to a three-dimensional spatio-temporal compression. And [14] combines wavelet with DTW for further energy savings.

There are some algorithms that can be used in both spatial and temporal domain. PINCO, a pipelined in-network compression scheme, is discussed in [15], whose basic idea is trading high data transmission latency for low transmission energy consumption with combining data packets into one packet. [16] achieves a lossless data compression by encoding information in the ordering of the sensor's remaining packets, which is known as "coding by ordering". A two-stage DPCM coding scheme implemented in [17] consists of temporal and spatial stages that compress data by making predictions based on samples from the past and helping sensors.

III. SPATIAL COMPRESSION DESIGN

Unlike the temporal compression algorithm, which is usually executed in sensor node singly and is no need for it to think of the way of routing or how the sensor nodes collaborate with each other, spatial compression schemes need to take into account the additional costs for communication. Therefore, spatial compression is considered to be a much more complicated scheme that combines routing and collaboration within compression.

A. Problem Setup

LTC, whose main idea is analogous to run-length encoding, makes an effort to represent a long sequence of similar data with a single symbol. Being a lossy compression with tunable accuracy, LTC is suitable for the dataset which is largely continuous and very slowly changing. By right of its remarkable compression performance and low computational complexity, LTC becomes a commonly used method in time domain.

Since measurements sampled from different nodes which are located close to each other have the same characteristics as the datasets collected in subsequent time instances, our work focuses on extending its application to the spatial domain. In order to avoid excessive energy losses in communication, the process of node interaction needs to be designed carefully.

B. Improved Scheme for Spatial Compression

For the sake of simplicity, we consider a one-dimensional network model for our first research, which can be easily expanded into other complex topology, for example, tree network. As showed in Figure 1, sensor nodes denoted as 1, 2, 3...N and the sink (node N+1) are placed equidistantly on a line. The whole set of measurements at the network nodes is denoted as a vector $X=(X_1, X_2, \dots, X_N)$, where X_i corresponds to the sample obtained by node i . Thus, X represents a space-sampled version of the original data with sampling step d along the line.



Figure 1. A one-dimensional network model

As mentioned above, the basic idea of the data compression is searching a linear trend in data stream. Since the datasets are dispersed in different nodes, additional costs for communication are inevitable. To minimize the energy consumptions of the whole network, we propose to adopt a pipelined compression scheme and introduce partial computation during spatial compression.

The data stream transmits along the one-dimensional network can be classified into two types: one is denoted as *trans_com*, which cannot be changed except for adding new content into it; the other is denoted as *trans_par*, which means partial computational results and will be modified subsequently.

For the different situations encountered in various datasets, we classify them into three cases and the scheme we propose is as follows:

1) Initialization: At the beginning of the scheme, the first two samples obtained by node1 and node2 are transmitted to node3 with X_1 contained in *trans_com* and ΔX_1 contained in *trans_par*, where ΔX_1 is obtained by $(X_2 - X_1)$, namely the difference between X_2 and X_1 , and then go to **Case 1** for next node.

2) Case 1: When current node receives the information from previous one, it adds its sample with the form of ΔX_2 , which is regarded as the difference between the current node's sample and the previous node's, into *trans_par*. Until now, there are three samples in the current node (one in *trans_com* and two in *trans_par*). Based on the error bound e decided by applications, two sector regions can be constructed. The current node estimates whether the two regions are overlapping. If not, add ΔX_1 into *trans_com*, change *trans_par*

with ΔX_2 to be ΔX_1 and repeat **Case 1** to the next node. Else, calculate the overlap region, change *trans_par* into the form of $\langle ul, ll, rl \rangle$, which corresponds to *highline*, *lowline* and *duration*, respectively, and go to **Case 2** for next node.

3) Case 2: Add the new sample of the current node to make a new sector region. Estimate whether there is overlap according to *trans_par*. If not, add $\langle k, rl \rangle$ into *trans_com*, where k is considered as the slope of the line after compression, and then go to **Case 1** for next node. Else, update *trans_par* and repeat **Case 2** for next node.

The compression process finishes once *trans_com* and *trans_par* arrive at the sink node. Recover data is taken at the sink and the error between the reconstruction data and the raw samples is bounded with the margin of error ϵ defined before compression.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate our algorithm by comparing with using a 5/3-wavelet transformation proposed in [11] and without any compression. The datasets obtained from the Tropical Atmosphere Ocean project (TAO, <http://www.pmel.noaa.gov/tao/>) contain various types of environmental data, such as temperature, relative humidity, salinity, pressure, winds and so on. Without loss of generality, we choose two datasets among them: relative humidity and temperature, which represent the data types of fast change and slow change, correspondingly. The nodes we select are deployed over the area with the total number of 50.

The total energy costs (E_{total}) include three parts: transmission (E_{TX}), reception (E_{RX}) and compression (E_{MCU}). We formulate them as follows:

$$E_{TX} = \sum_{i=1}^N P_{TX}(d) * L_i * T_{tran} \quad (1)$$

$$E_{RX} = \sum_{i=2}^N P_{RX} * L_i * T_{tran} \quad (2)$$

$$E_{MCU} = \sum_{i=1}^N P_{MCU} * T_i \quad (3)$$

$$E_{total} = E_{TX} + E_{RX} + E_{MCU} \quad (4)$$

where L_i denotes the compressed data size in node i , and T_{tran} is the time cost for transmitting one byte of data. The transmit power P_{TX} is closely related to the communication distance d , while the MCU power P_{MCU} and the received power P_{RX} are approximately constant when the related modules work in the active mode (like MicaZ in our tests). T_i is the time cost for compressing data in node i . Furthermore, we use uniform quantization and no entropy coding in the simulations. And the level of wavelet transformation we take is one for an analogous computational complexity and information interaction to our spatial compression.

Figure 2 shows the percent savings in bits resulting from compressing relative humidity and temperature with a function of error using wavelet transformation and our scheme, where bits is denoted as the amount of bits received in the sink and the savings are obtained by compare the amount to the

original one. Error is described as the root-mean-square differences between the raw data and the reconstruction one, which is represented as an error bound determined by each application. In this simulation, we take $\pm 0.1\%$ and $\pm 0.5^\circ\text{C}$ corresponding to relative humidity and temperature, respectively.

Figure 3 shows the percent savings in energy resulting from compressing with the same function of error, where energy is calculated by the models of energy consumption mentioned above and involves the total energy costs for transmitting the whole samples to the sink.

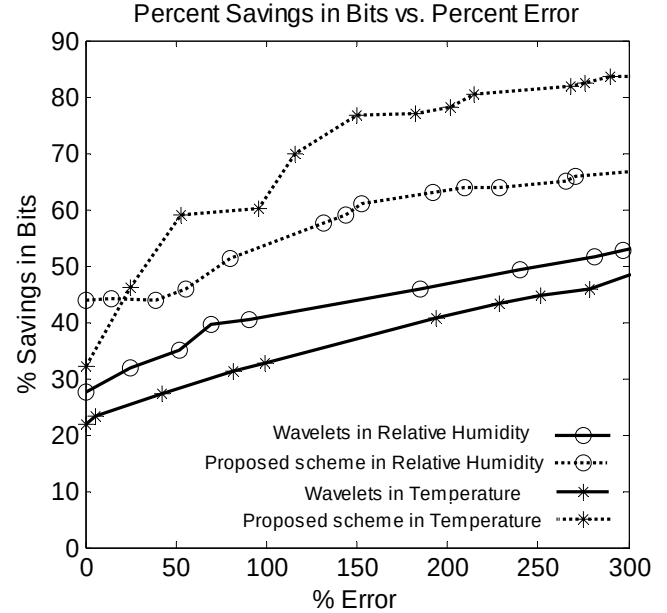


Figure 2. Savings in bits as a function of error

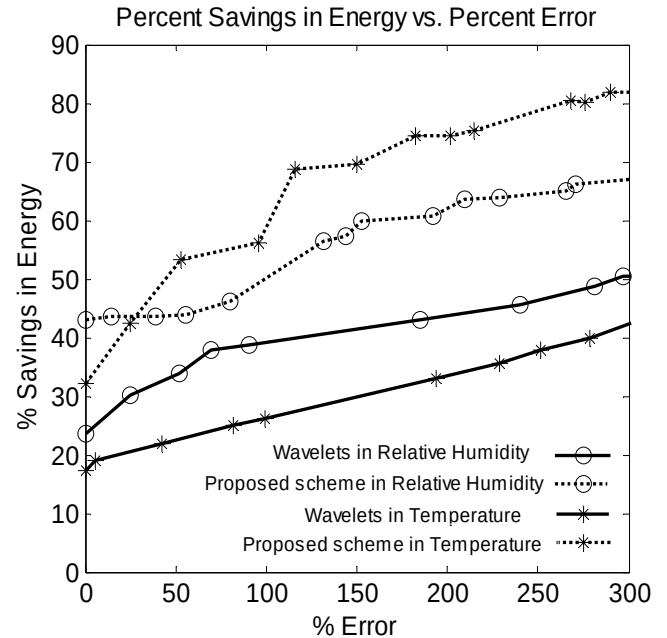


Figure 3. Savings in energy as a function of error

The simulation results show that wavelet achieves 40% savings in bits and 40% savings in energy when error is $\pm 0.1\%$ in relative humidity, while the proposed scheme achieves 55% and 52%, respectively. In datasets of temperature, wavelet obtains 32% savings in bits and 25% savings in energy when error is $\pm 0.5^\circ\text{C}$, while the proposed scheme obtains 60% and 59%, respectively, which is twice over the results in wavelet. More savings can be obtained with the increasing of error and the savings due to compression would approach 100% ultimately. In addition, when distortion of data is not allowed during compression, the proposed scheme will alter to differential coding.

V. CONCLUSIONS

A lightweight compression algorithm is proposed in this paper, which can be used for raw data with spatial correlation. By adopting pipelining and introducing the concept of partial computation, we can decorrelate the original data with little inter-sensor communication over short distances. The energy consumption of the proposed scheme can reduce to half of the original cost, while still keeping a satisfying data accuracy. Compared with the wavelet compression, our scheme achieves more savings both in the amount of bits and energy usages under a comparable computational complexity.

Several extensions of the proposed scheme are worth further investigation. We will extend the schemes for spatio-temporal compression and integrate it with the other protocols implemented in wireless sensor networks. Extension of the applicability to a two-dimensional network topology is our current work.

ACKNOWLEDGMENT

In this paper, the research is sponsored by the Nature Science Foundation of Zhejiang Province (Project No. LQ13F010005) and Ningbo Natural Science Foundation (Project No. 2011A610185). The work is supported by Ningbo Sc. & Tech. plan project (2014B82015).

REFERENCES

- [1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, pp. 102-114, August 2002.
- [2] T. Schoellhammer, E. Osterweil, B. Greenstein, et al, "Lightweight temporal compression of microclimate datasets," *Proceedings of the 29th Annual IEEE International Conference on Local Computer Networks IEEE Computer Society*, 2004, pp. 516-524.
- [3] D. Estrin, M. Srivastava, A. Sayeed, "Tutorial on wireless sensor networks," *Technologies Protocols & Applications*, vol. 13, pp. 317-328, 2002.
- [4] Y. Liang, "Efficient temporal compression in wireless sensor networks," *IEEE 36th Conference on Local Computer Networks*, 2011, pp. 466-474.
- [5] J. M. Zhang, Y. P. Lin, S. W. Zhou, et al, "Haar wavelet data compression algorithm with error bound for wireless sensor networks," *Journal of Software*, vol. 21, pp. 1364-1377, 2010.
- [6] Y. Mo, Y. Qiu, J. Liu, et al, "A data compression algorithm based on adaptive Huffman code for wireless sensor networks," *IEEE 4th International Conference on Intelligent Computation Technology and Automation*, 2011, pp. 3-6.
- [7] S. S. Pradhan, J. Kusuma, K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, pp. 51-60, 2002.

- [8] G. Hua, C. W. Chen, "Distributed source coding in wireless sensor networks," *IEEE 2nd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks*, 2005, pp. 1-6.
- [9] J. Chen, X. Han, "The distributed source coding method research based on clustering wireless sensor networks," *International Journal of Sensor Networks*, vol. 17, 2015.
- [10] C. Braccini, F. Davoli, M. Marchese, et al, "Surveying multi-disciplinary aspects in real-time distributed coding for wireless sensor networks," *Sensors*, vol. 15, pp. 2737-2762, 2015.
- [11] A. Ciancio, A. Ortega, "A distributed wavelet compression algorithm for wireless multihop sensor networks using lifting," *IEEE International Conference on Acoustics, Speech, and Signal*, 2005, pp. 825-828.
- [12] H. Dong, J. Lu, Y. Sun, "Adaptive distributed compression algorithm for wireless sensor networks," *IEEE International Conference on Innovative Computing, Information and Control*, 2006, pp. 283-286.
- [13] C. Roadknight, L. Parrott, N. Boyd, et al, "A layered approach to in situ data management on a wireless sensor network," *Proceedings of the Intelligent Sensors, Sensor Networks and Information Processing*, 2004, pp. 85-90.
- [14] S. W. Zhou, L. I. Lan, "DTW-based multi-wavelet data compression algorithm for wireless sensor networks," *Journal on Communications*, vol. 43, pp. 121-124, 2014.
- [15] T. Arici, B. Gedik, Y. Altunbasak, et al, "PINCO: a pipelined in-network compression scheme for data collection in wireless sensor networks," *Proceedings of the 12th International Conference on Computer Communications and Networks*, 2003, pp. 539-544.
- [16] D. Petrovic, R. C. Shah, K. Ramchandran, et al, "Data funneling: routing with aggregation and compression for wireless sensor networks," *Proceedings of the 1st Sensor Network Protocols and Applications*, 2003, pp. 156-162.
- [17] H. Luo, Y. C. Tong, G. Pottie, "A two-stage DPCM scheme for wireless sensor networks," *International Conference on Acoustics, Speech, and Signal Processing*, 2005, pp. 661-664.



Beihua Ying received her Ph.D. degree from Tsinghua University, Beijing, China, in 2010 and the BS degree from Xidian University, Xi'an, China, in 2004, both in Electronic Engineering. She is currently a lecturer with the school of Information Science and Engineering, Ningbo Institute of Technology, Zhejiang University. Her research interests include design and analysis of energy-efficient data process for wireless sensor networks, data gathering and routing protocol in wireless ad hoc and sensor networks, and optimization problems.