International Journal of Multimedia and Ubiquitous Engineering
Vol. 9, No. 11 (2014), pp. 21-30
http://dx.doi.org/10.14257/ijmue.2014.9.11.03

An Efficient Compression Method in Wireless Sensor Networks

Byoungyup Lee[1], Myoungho Yeo[2], Kyungsoo Bok[2] and Jaesoo Yoo[2*]

[1]*Paichai University,*
[2]*Chungbuk National University, Korea*
bylee@pcu.ac.kr, myungho.yeo@gmail.com, {ksbok, yjs}@chungbuk.ac.kr

Abstract

Sensor data exhibit strong correlation in both space and time. Many algorithms have been proposed to utilize these characteristics. However, each sensor just utilizes neighboring information, because its communication range is restrained. Information that includes the distribution and characteristics of whole sensor data provides other opportunities to enhance the compression technique. In this paper, we propose an orthogonal approach for compressing sensor readings based on a novel feedback technique. That is, the base station or a super node generates Huffman code for the compression of sensor data and broadcasts it into sensor networks as Huffman code. All sensor nodes that have received the information compress their sensor data and transmit them to the base station. We call this approach as feedback-diffusion and this modified Huffman coding as sHuffman coding. In order to show the superiority of our approach, we compare it with the existing data compression algorithms in terms of the lifetime of the sensor network. As a result, our experimental results show that the whole network lifetime was prolonged by about 30%.

Keywords: *Sensor network, data compression, Huffman code, network lifetime, suppression*

1. Introduction

Wireless sensor networks become in the limelight for the variety of RFID applications like environment monitoring, smart spaces, medical applications, and precision agriculture. The sensor networks collect useful information such as temperature, humidity and seismic intensity. They transmit sensor readings to the base station for sophisticated processes. Because sensor nodes have limited batteries and consume a lot of energy for communication, energy-efficient methods are required to reduce the network traffic[1-4].

Data compression techniques are traditional and effective methods to reduce the network traffic. Generally, sensor readings are correlated by both space and time. Recently, many researches have been proposed to compress sensor readings with data correlations. [5] proposes a temporal suppression scheme. It assumes error bound of sensor readings. The basic idea is that sending sensor readings is suppressed if they are bounded in error bound of the latest reported data. Then, the base station regards current data as the latest reported data with temporal correlation.

[6] has proposed a data suppression algorithm using a spatial correlation. All sensor nodes receive and transmit their sensor readings according to their corresponding time slot. Sensor nodes overhear the transmitted data from neighbors to the base station while waiting for their own time slots. At this time, each sensor calculates an average of overheard data and compares the average with its own reading. If the reading is equal to the average, it doesn't transmit its own reading. In [7], sensor nodes organize cluster forms. Member nodes in each cluster transmit their own readings to their cluster head.

* Corresponding Author

ISSN: 1975-0080 IJMUE

The cluster head compresses collected data and removes duplicated readings. [4] has proposed a clustering algorithm based on data correlation. It improves the performance of compression by well organized clusters. However, conventional algorithms exploit just local distribution like historical data or neighbors in the range of communication.

In this paper, we propose a feedback diffusion algorithm to utilize global distribution of sensor data. The base station determines global distribution from collected data in the sensor network and broadcasts the Huffman code to sensor nodes. The sensor nodes compress their own data using the Huffman code and transmit them to the base station. For more efficient data compression by feedback distribution, it is important to reduce communication costs and improve the compression rate. Thus, we propose a variant of the Huffman coding, called *sHuffman*, for sensor networks. In order to show the superiority of our approach, we compare our proposed algorithm with existing algorithms through various simulations. Our experimental results show that the whole network lifetime was prolonged by about 30% and whole sensor data were compressed by about 40%.

The rest of this paper is organized as follows. In Section 2, we describe an existing data compression algorithm and the characteristic of Huffman coding as related works. In Section 3, we propose and mathematically analyze the feedback diffusion algorithm. In Section 4, we describe the Huffman coding for sensor networks and its procedure in detail. In Section 5, we show the superiority of our proposed compression algorithm via performance evaluation and analysis. Finally, we conclude in Section 6.

2. Related Work

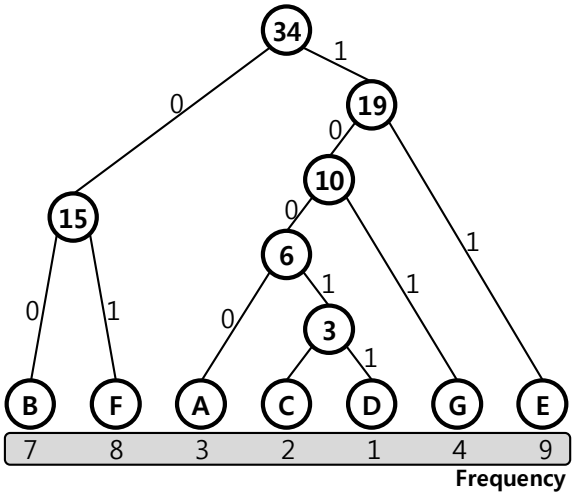
2.1. Sensor Data Compression Algorithm

Sensor readings are strongly correlated by both space and time [4]. There are many approaches with these correlations such as spatial compression and temporal compression. Temporal compression algorithms transmit sensor readings to the base station if current data has a change as compared with the latest reported data. The base station regards unreported data as the latest reported data. [5] proposes another temporal compression algorithm with error rate. If the change of readings is bigger than an allowed error rate, sensor nodes transmit their readings to the base station.

Data compression algorithm utilizing a spatial feature utilizes the similarity of data of the neighbor nodes. [6] proposed data compression to utilize the mean operator. All nodes receive different time slots and transmit their own collected data according to the received time slot. Sensor nodes overhear data that neighbor nodes transmit the data to the base station while waiting for their time slots. At this time, each sensor node calculates an average of whole overheard data. If this average is equal to its own collected data, the collected data is not transmitted to the base station. To improve the data compression efficiency, in-network processing methods such as the clustering and the tree structure were proposed [7-8]. In cases that the data of a neighbor node have a high correlation, transmitting merged data at in-network is more efficient than transmitting data at each node to the base station. In [7], sensor nodes make up clusters and each node of the cluster transmits its own collected data to the head node. At this time, the head node of the cluster compresses data by removing duplicated data.

2.2. Huffman Coding

Huffman coding is the lossless compression algorithm as one of the entropy encoding algorithms. It uses a variable length code according to the occurrence frequency of symbols as shown in Figure 1(a). The tree level is determined by the weight of each symbol. As shown in Figure 1(b), bit patterns that are along the path from the root node to the specific node are allocated to each symbol. Therefore, short bits are allocated to high frequent symbols and long bits are allocated to the relatively low frequent symbols.



(a)

Symbol	Huffman code
A	1000
B	00
C	10010
D	10011
E	11
F	01
G	101

(b)

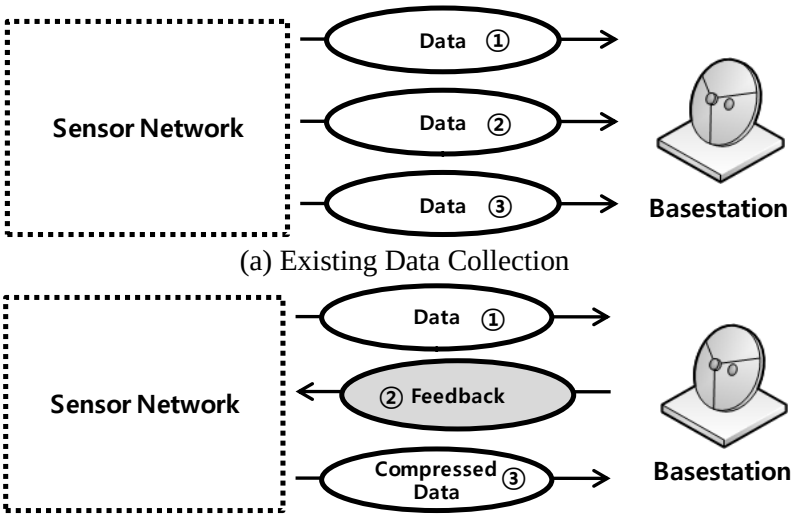
Figure 1. Huffman Coding

3. Feedback-Diffusion Algorithm

In existing algorithms, the base station collects just data from the network. Figure 2(a) shows the existing sensor data collection. As we mentioned above, the existing sensor data compression techniques utilize the historical data of each node or the data of neighbor nodes. In this paper, we propose a novel data compression scheme utilizing an interaction between the base station and sensor networks. In other words, data compression is done according to each algorithm in the sensor networks and the compressed data are transmitted to the base station. Figure 2(b) represents that each node in the sensor network efficiently compresses its own data by utilizing redistributed data which is extracted from initially collected global data. We define the redistributed data as feedback. Our proposed scheme has an advantage which reflects global distribution.

We suppose that there is a TAG-based sensor network which consists of N nodes. Each sensor data v is compressed into v' by compression algorithm. The equation (1) indicates the communication costs of data transmitted to the base station. In the equation (1), avgDist is a mean distance from each sensor node to the base station. Thus, each sensor node transmits data which have a log(v') bit size at the each round.

$$C_{tag} = N \cdot avgDist \cdot log(v') \qquad (1)$$



(a) Existing Data Collection

(b) Data Collection with Feedback Diffusion

Figure 2. An Example of Feedback Diffusion

Equation (2) denotes the communication costs of the proposed feedback diffusion algorithm. The communication cost C_f is composed of the sum of data collection cost C_{tag} based on TAG, feedback diffusion cost C_{fd} and reception costs C_{fc} of compressed data by feedback *F*. T_c means the number of data collections and T_f means the number of broadcasting Huffman code.

$$\begin{aligned} C_{fd} &= N_{nl} \cdot log(F) \\ C_{fc} &= N \cdot avgDist \cdot log(v'') \\ C_f &= \frac{C_{tag} + T_f \cdot C_{fd} + T_c \cdot C_{fc}}{T_c} \end{aligned} \quad (2)$$

Equation (3) denotes the calculation procedure of communication costs to design an efficient feedback diffusion scheme. To design the feedback diffusion scheme efficiently, we have to improve a collection gain G_{fc} of the compression scheme and decrease a feedback diffusion cost C'_{fd}. Since the original costs of feedback diffusion are proportional to F and T_f, we study a scheme to reduce a feedback size and the number of feedback diffusion distributions.

$$\begin{gathered} C_{tag} > C_f \\ T_c \cdot C_{tag} > C_{tag} + T_f \cdot C_{fd} + T_c \cdot C_{fc} \\ T_c \cdot (C_{tag} - C_{fc}) - (C_{tag} + T_f \cdot C_{fd}) > 0 \\ G_{fc}(T_c) = T_c \cdot (C_{tag} - C_{fc}) \\ C'_{fd}(T_f) = (C_{tag} + T_f \cdot C_{fd}) \\ G_{fc}(T_c) - C'_{fd}(T_f, F) > 0 \\ C'_{fd}(T_f, F) \propto T_f, F \end{gathered} \quad (3)$$

4. Huffman Coding in WSNs

In this section, we describe an algorithm to reduce the size of Huffman code and the number of feedback diffusion. Our proposed feedback diffusion algorithm reduces data size by representing the sensor data to the bit patterns. Huffman coding is a representative algorithm to efficiently express data with the minimum bits. It is difficult to directly apply the Huffman coding as the Huffman code to sensor networks. The reason is that diffusing

Huffman codes needs a lot of communication costs. In this paper, we first analyze the features of sensor data. Then we propose the sensor network version of the Huffman coding algorithm that reduces the size of the Huffman code.

4.1. Creation of Huffman Code

In this paper, we propose a feedback diffusion algorithm based on data frequency. Generally, sensor data exhibit strong correlation in both space and time. Therefore, there is a chance to generate specific data with high frequency. Thus, we can improve compression efficiency by expressing high frequent data as bit patterns. However, diffusing the Huffman codes as the Huffman code leads a lot of energy consumption. To overcome this problem, we generate a light weight feedback by utilizing the features of the sensor data.

Feature (1). Sensor data have some error bound.

The error bound of sensor data can be predefined differently for many applications. Sensor data are grouped according to error bound ε as shown in Equation (4). v_real and v_base mean real sensor data and base data for grouping respectively. We reduce the number of sensor data considered for grouping them.

$$|v_{real} - v_{base}| \leq \varepsilon \qquad (4)$$

Feature (2). Some sensor nodes sense completely wrong data.

Sensor nodes can get some errors easily, because they are deployed in unstable environments. Therefore, some of them may transmit absolutely wrong data to the base station. We define threshold fcutoff to exclude wrong sensor readings. Figure 3 shows a procedure for creating Huffman code by sHuffman. First, sensor data are grouped within error bound ε. Next, data groups which have lower frequency than threshold fcutoff are pruned. Finally, the lighten Huffman code (called as sHuffman) is organized and is broadcasted into sensor nodes.

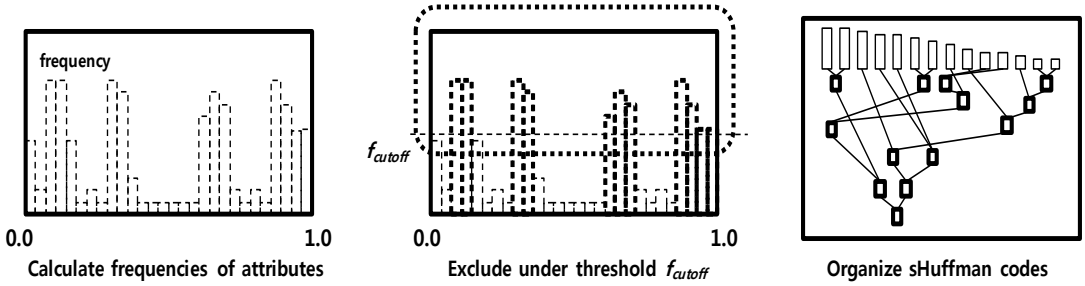


Figure 3. A Procedure for Creating sHuffman Codes

4.2. Sensor Data Compression Utilizing a Feedback

Each sensor node attempts to quantize its reading with sHuffman codes. Figure 4 shows an example of gathering sensor readings with the feedback diffusion. First, the base station generates feedback and diffuses it to the sensor network. Each sensor node stores this feedback. If current sensor readings are matched to the feedback within the error bound(ε), they can be expressed as corresponding sHuffman codes. Otherwise, we have two methods. One method is to ignore them because they may be errors by Feature (2). Another method is to collect them as uncompressed data directly. Some applications want to improve the accuracy of data gathering. Then, we can collect sensor readings as the combination of sHuffman codes and raw data. Although the communication cost is increased, the communication cost is not high because sensor readings which are only transported as raw data have low frequency relatively.

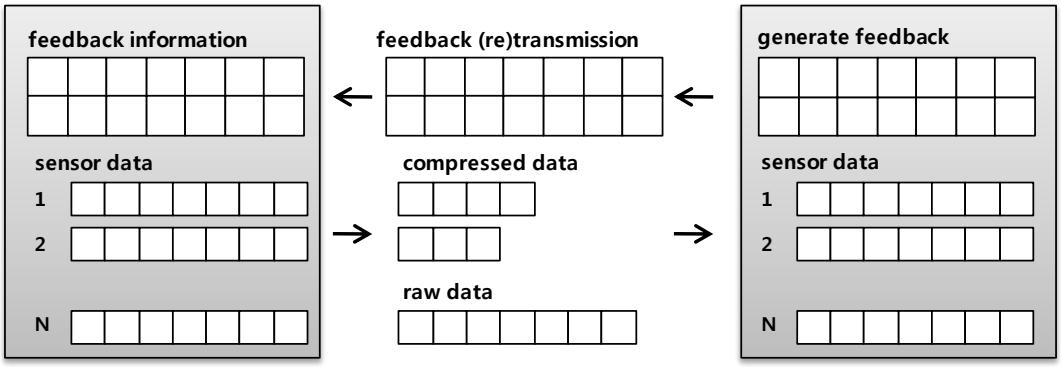


Figure 4. Gathering Sensor Readings with the Feedback Diffusion

4.3 Update Huffman Code

The Huffman code is generated by utilizing the collected sensor data at a specific time. Sensor readings are compressed based on the diffused feedback. However, the compression-efficiency is decreased because the difference of current sensor readings and the Huffman code is increased. If we diffuse the feedback to improve compression rate at every round, sensor nodes get recent Huffman code, but consume a lot of energies. On the other hand, if we do not update the feedback, the compression efficiency is degraded and more energy is consumed. To overcome this problem, we have to keep the number of collected data to determine when the feedback must be updated. We define a collected degree of quantized data as the hit rate. If the hit rate is lower than the predefined hit rate, the feedback is updated.

5. Performance Evaluation

5.1 Experiment Environment

To show the superiority of our approach, we compare it with existing data compression algorithms in terms of the network lifetime and network traffic. Table 1 indicates environment parameters. The consumption energy model to transmit a message is defined as {message size}·[{transmit cost} + {amplification cost}·{distance}]. Transmission cost is 50nJ/b and amplification cost is 100pJ/b/m2. The consumption energy model to receive a message is {message size}· {receive cost} and reception cost is set as 50nJ/b[10]. To guarantee the correctness of sensor data, data that is not covered by the feedback are transmitted to the base station as a raw data.

Table 1. Performance Evaluation Parameter

parameter	change	default
Tree level	2~10	3
Sensor identifier	-	4byte
Size of sensor reading	-	4byte
ε	0~4%	1%
Hit rate	10~95%	80%
f_{cutoff}	1~16%	4%

Figure 5 shows the network topology for simulations. Figure 6 shows the representative segments of the TEMP data traces in the simulation. The sensor readings are simulated using the real traces provided by the Live from Earth and Mars (LEM) project - at the University of Washington [9]. We extracted many subtraces starting at

different times. We assigned each subtrace to the reading of each sensor. We controlled the variance of the readings with the starting time of subtraces.

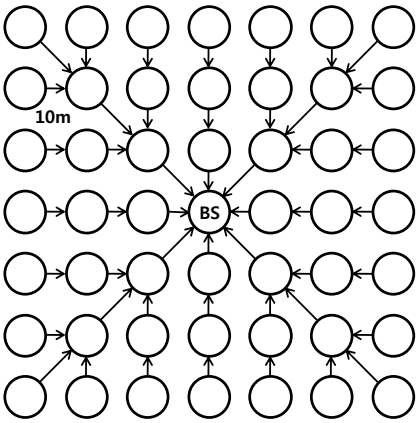


Figure 5. TAG-based Network Topology

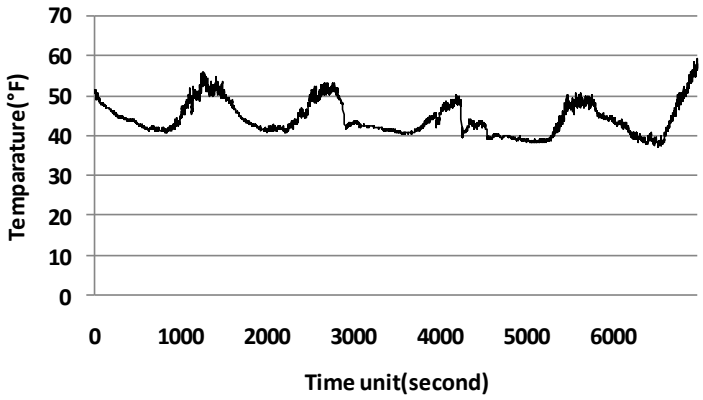


Figure 6. Data Model Used in Simulation

5.2 Network Traffic

We compare the size of feedback generated by our proposed scheme with that of Naïve feedback scheme which ex presses sensor readings as bits with fixed length. Figure 7 shows the size of the Huffman code. Our proposed scheme is more scalable than the Naïve feedback scheme, because sensor readings are grouped and sensor readings with low frequency are excluded from the feedback. The size of the feedback for the Naïve feedback scheme is rapidly increased, because the number of generated samples is proportionally increased as the size of network grows.

Figure 8 presents the average network traffic as the number of sensor nodes increases. Our proposed scheme reduces the network traffic by about 40%. This result is due to Huffman coding which allocates a code according to the frequency of data.

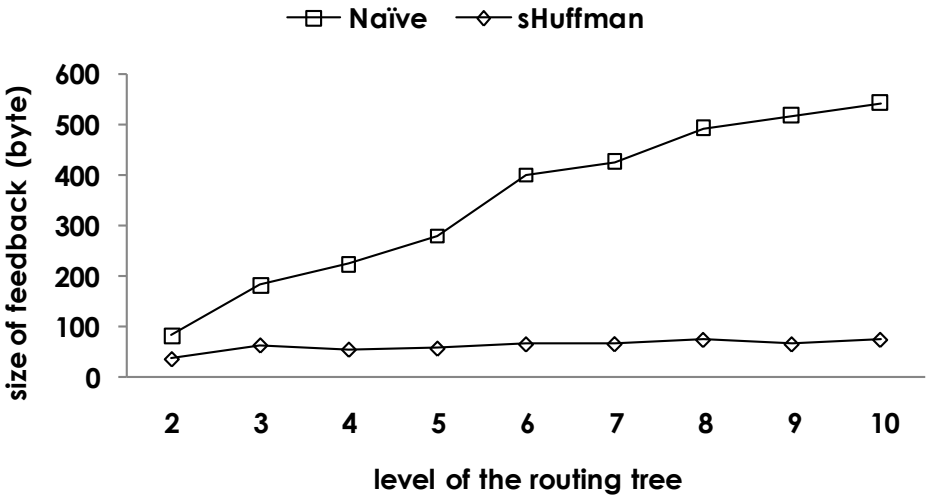


Figure 7. Size of Feedback

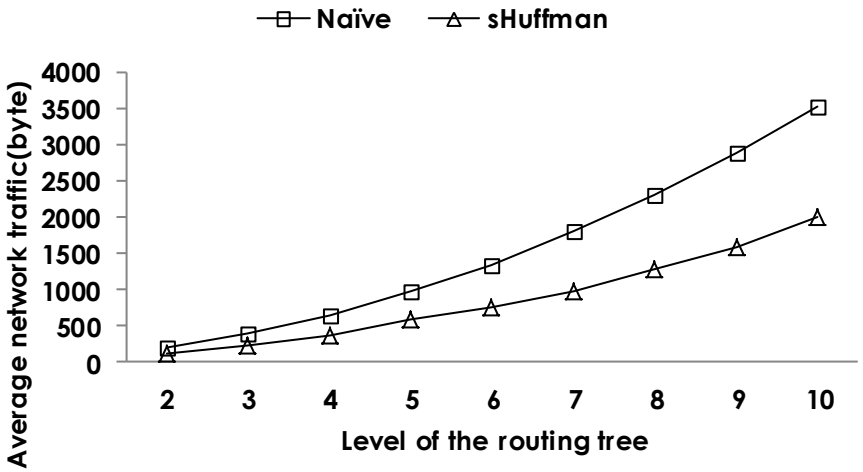


Figure 8. Average Network Traffic

5.3. Network Lifetime

Figure 9 shows the network life with or without conventional compression algorithms. FM(Flooding Method) is an aggregation algorithm based on TAG. To evaluate - the existing compression algorithms, [5] is applied. In the result, the network lifetime of our proposed algorithm is prolonged by about 30%. In case of our proposed algorithm with the existing compression algorithm, it shows additional energy-efficiency.

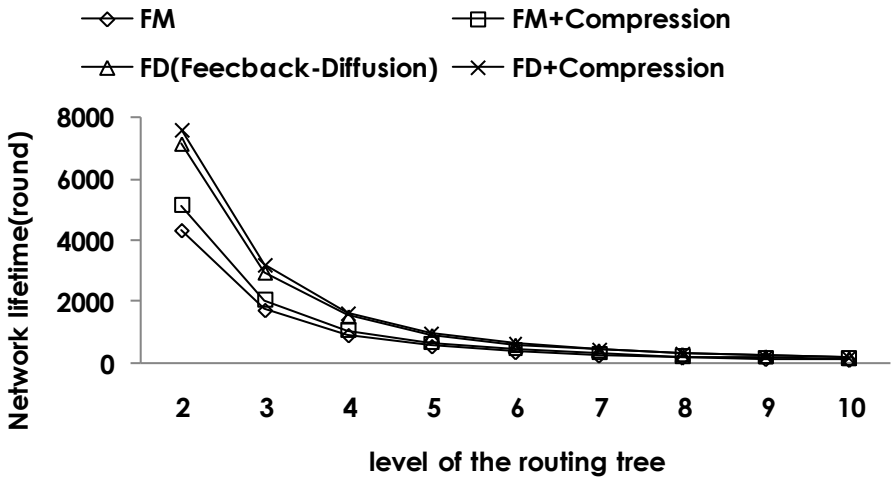


Figure 9. Network Lifetime

6. Conclusions

In this paper, we have proposed a feedback diffusion algorithm based on the variant of Huffman coding, called sHuffman. While existing algorithms just exploited local distribution, our proposed scheme compresses sensor readings efficiently from global distribution. In order to show the superiority of our approach, we compared it with the existing aggregation algorithms in terms of the lifetime of the sensor network. As a result, our experimental results have shown that the whole network lifetime was prolonged by about 30% and we can improve energy efficiency by utilizing existing compression algorithms in parallel. Also, we confirmed that it is possible to efficiently collect data by utilizing an error of the sensor network and an error bound of the sensor data. In the future, we will apply our algorithm to the real sensor network applications.

Acknowledgements

This work was supported by the ICT R&D program of MSIP/IITP. [14-824-09-001, Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis , by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIP) (No.2013R1A2A2A01015710) and by the Ministry of Education(MOE) and National Research Foundation of Korea(NRF) through the Human Resource Training Project for Regional Innovation(no. 2013H1B8A2032298).

References

[1] D. Estrin, L. Girod, G. Pottie and M. Srivastava, "Instrumenting the World with Wireless Sensor Networks", In Proceedings of International Conference Acoustics, Speech, and Signal Processing, **(2001)**.
[2] G. J. Pottie and W. J. Kaiser, "Wireless Integrated Network Sensors", In Proceedings of Comm. ACM, **(2000)**, pp. 51-58.
[3] I. F. Akyildiz, W. Su, Y. Sankarasubramanism and E. Cayirci, "A Survey on Sensor Networks", In Proceedings of IEEE Communications Magazine, **(2002)**.
[4] M. H. Yeo, M. S. Lee, S. J. Lee and J. S. Yoo, "Data Correlation-Based Clustering Algorithm in Wireless Sensor Networks", The 2008 International Symposium on Computer Science and its Applications, **(2008)**.
[5] M. Sharaf, J. Beaver, A. Labrinidis and P. Chryanthis, "Tina: A scheme for temporal coherency-aware in-network aggregation", In Proceedings of the ACM Workshop on Data Engineering for Wireless and mobile Access, **(2003)**.
[6] X. Meng, L. Li, T. Nandagopal and S. Lu, "Event contour: An efficient and robust mechanism for tasks in sen-sor networks", In Proceedings of Technical report, **(2004)**.

[7] S. Pattem, B. Krishnamachari and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks", In Proceedings of International Conference on Information Processing in Sensor Net-works, **(2004)**.
[8] D. Petrovic, R. Shah, K. Ramchandran and J. Rabaey, "Data funneling: Routing with aggregation and compression for wireless sensor networks", In Proceedings of the 2003 IEEE Sensor Network Protocols and Applications, **(2003)**.
[9] Live from Earth and Mars (LEM) Project, http://www-k12.atmos.washington.edu/k12/grayskies/, **(2006)**.
[10] W. Heinzelman, "Application-Specific Protocol Architectures for Wireless Networks", PhD dissertation, Massachusetts Inst. Of Technology, **(2000)**.

