

# Multi Data Fusion Model Based on Information Entropy in Sensor Network

Zhao Li-ming<sup>1,2</sup>, Liu He-ping<sup>1</sup>

1.School of Automation and Electrical Engineering, University of Science and Technology, Beijing ,  
China

2.Institute of Information, Guangdong Ocean University, Zhanjiang, China

**Keywords:** Wireless Sensor Network; Data Fusion; Compressed Sensing; Mixed Integer Programming; Greedy Algorithm; Network Life Cycle

**Abstract.** Data fusion is one of the research hotspots in the field of wireless sensor network. In allusion to the problem of excessive energy consumption of existing fusion methods, a data fusion scheme with minimized energy consumption is proposed in this article on the basis of the compressed sensing theory. Firstly, the influence of different fusion modes on data collection performance is analyzed; then, the influence of route and mixed CS fusion on energy optimization is considered in order to model the data fusion problem into the mixed integer programming problem with minimized energy consumption, and meanwhile a greedy algorithm based on fusion node increase is proposed to solve the above problem, wherein all nodes in the network are divided into fusion nodes used for CS encoding and forwarding nodes used for directly forwarding their own data and the received data; finally, a data fusion tree with minimized energy consumption is obtained. The result of the simulation experiment shows that the scheme proposed in this article is effective and superior to the traditional methods in the aspects of data recovery accuracy, network life cycle, etc.

## Introduction

As a research hotspot widely concerned at home and abroad in recent years, wireless sensor network (WSN) is usually deployed in severe environments or the environments difficult for human to enter. There are many nodes distributed in a wireless sensor network, and these nodes are usually transported by users through airplanes or vehicles and then randomly put into designated areas to execute complex tasks. Since all nodes have limited storage space and computing ability and meanwhile the communication range is also limited, these nodes must rapidly and independently cooperate with each other in order to efficiently finish the tasks in a cooperative working manner[1].

Data fusion[2] is not only an important technology in WSN, but also a research hotspot at present. This technology can be used to process mass original data collected by the sensor nodes in the network through certain algorithm, remove the redundant information contained therein and only transmit a small amount of useful processing results to the aggregation nodes. The adoption of data fusion technology can significantly reduce the volume of the data transmitted in WSN, reduce data collision and relieve network congestion, thus effectively reducing energy consumption and prolonging network life. Due to the limitation of the physical environment and the energy or capacity of the sensor nodes, the existing data fusion schemes involve in large energy consumption and the data fusion result is not accurate when relevant nodes suffer from fault. Therefore, this article is focused on researching the data fusion scheme for wireless sensor network, and the successful implementation of the scheme has important theoretical significance and application value for ensuring the reliability of the wireless sensor network.

## System Model

### Network Scene

The data fusion problem in wireless sensor network is researched in this article and is based on the following assumptions:

$n$  sensor nodes ( $s=1,2,\dots,n$ ) are randomly distributed in a square zone  $A$  with the area of  $M \times M$ . The whole wireless sensor network forms an interconnected undirected graph  $G(V, E)$ , wherein  $V$  is the sensor node set,  $E$  is the edge set in  $G$ , and special nodes  $s \in V$  are the sink nodes responsible for collecting the information from the whole network.  $n$  and  $\mathcal{V}$  respectively denote the base numbers of  $V$  and  $E$ . If two sensor nodes  $V_i$  and  $V_j$  are located within the communication radius of the opposite side, then  $(V_i, V_j) \in E$ . In order to realize algorithm extendibility, there is no need for the nodes to know their location information, and the nodes only need to easily exchange one "Hello" message with their neighbors to know their neighbor information. Additionally, the network has following characteristics:

1) The network is an interconnected static network, and the nodes are not moved after deployment;

2) The energy consumption of an optional node  $i$  in a wireless sensor network can be expressed as follows:  $E_i = E_{i-MCU} + E_{i-TCR} + E_{i-SB}$ , wherein  $E_{i-MCU}$  denotes the energy consumption of the micro-control unit at node  $i$ ;  $E_{i-SB}$  denotes the computation energy consumption of the CPU at node  $i$ ;  $E_{i-TCR}$  denotes the communication energy consumption for node  $i$  to receive and transmit data.  $E_{i-TCR}$  can be further expressed as follows:  $E_{i-TCR} = E_{i-TCR-RX} + E_{i-TCR-TX(d_i)}$ , wherein  $E_{i-TCR-RX}$  denotes the energy consumption for node  $i$  to receive data;  $E_{i-TCR-TX(d_i)}$  denotes the energy consumption for node  $i$  to transmit data and is a function related to distance  $d_i$ , and if the distance between data transmitting node  $i$  and the receiving node is larger, then more energy will be consumed. Under the assumption that the transmission route attenuation factor of the node in the network is 2, the total energy consumption of the wireless sensor network with  $N$  nodes can be expressed as follows[9]:

$$E_{total}(N) = C_1 + C_2 + C_3 \sum_{i=1}^N d_i^2 \quad (1)$$

In the above formula,  $C_1$ ,  $C_2$  and  $C_3$  are constants.

### Compressed Sensing Theory

The compressed sensing theory[10] is mainly adopted in this article for data fusion. In order to facilitate subsequent description, the basic procedure for the signal processing based on the compressed sensing theory is given as follows:

A discrete signal  $x$  with the length of  $N$  is assumed and can be expressed as  $N \times 1$  column vector  $x(n) \in R^N$ . Particularly, signal  $x$  can be sparsely expressed as follows on a transform basis  $\Psi = [\Psi_1, \Psi_2, \dots, \Psi_N]$ :

$$x = \sum_{i=1}^N s_i \Psi_i = \Psi s \quad (2)$$

In the above formula,  $s$  is the coefficient sequence of signal  $x$  on  $\Psi$ . If  $s$  only includes  $k$  nonzero elements, then  $x$  is called to be sparse to  $k$ .

Nonlinear measurement can be adopted to obtain the measured value of signal  $x$ :

$$y = \Phi x = \Phi \Psi s = \Theta s \quad (3)$$

In the above formula,  $\Phi$  is  $M \times N$  projection matrix ( $M = N$ ). Since  $y$  is  $M \times 1$  vector and  $x$  is  $N \times 1$  column vector, so the recovery of  $x$  in  $y$  is an underdetermined problem without definite solution. Since  $x$  is sparse to  $k$ , according to the compressed sensing theory, if  $\Theta$  meets RIP principle or  $\Phi$  is irrelevant to  $\Psi$ , the original signal  $x$  can be accurately recovered through

solving the following  $l_1$  norm optimization problem:

$$s = \arg \min_s \|s\|_{l_1} \quad s.t. \quad y = \Phi \Psi s \quad (4)$$

Linear programming technology can be used to solve Problem (4), and the known solving methods mainly include OMP[11], StOMP[12] and CP[13], etc.

### Mixed CS Fusion Mechanism

The following instance is taken to analyze the influence of different data fusion mechanisms on the data collection in WSN. As shown in Figure 1, the wireless sensor network is assumed to have  $n$  nodes. If the non-fusion mechanism is adopted and  $n-1$  nodes respectively transmit one sample datum to the  $n$ th node, then the  $n$ th node will output  $n$  samples. At this moment, the node with more child nodes needs greater data volume for data forwarding, namely: the larger the load, the greater the energy consumption, and this node is easy to die. Therefore, such node is not applicable to the data collection in WSN; if the lossy fusion mechanism is adopted, then the  $n$ th node will output 1 sample. Although the communication overhead is reduced and the energy is saved, the information of some nodes is lost, thus seriously influencing data fusion quality.

In order to avoid the above defects, this article is focused on researching the data fusion technology based on the compressed sensing theory. When the compressed sensing technology is adopted for data fusion (abbreviated as CS fusion), if CS encoding is directly implemented for all nodes in the network, such fusion is called common CS fusion. Generally speaking, the direct application of CS encoding to all sensor nodes for data collection is not the best choice. Under such fusion mode, each node load has  $k$  data samples; the data volume of the node load is usually less than  $k$  in the initial transmission stage, so it will cause the node to transmit excessive unnecessary data samples and increase the communication overhead. Therefore, a good fusion mode shall be as follows: when the quantity of the data samples output by the fusion node is more than the data volume after CS fusion, then CS fusion shall be adopted, or else, the data shall not be fused but shall be directly forwarded. Such fusion mode is called mixed CS fusion in this article. The specific implementation is as follows: the non-fusion mode is adopted for all nodes in the initial stage to obtain the sensing data and meanwhile a fusion factor  $k$  is given; for an optional node  $i$ , if the volume of the data received by node  $i$  from all downstream nodes is more than or equal to  $k$ , then node  $i$  shall be converted to adopt CS encoding for the fusion of the original data.

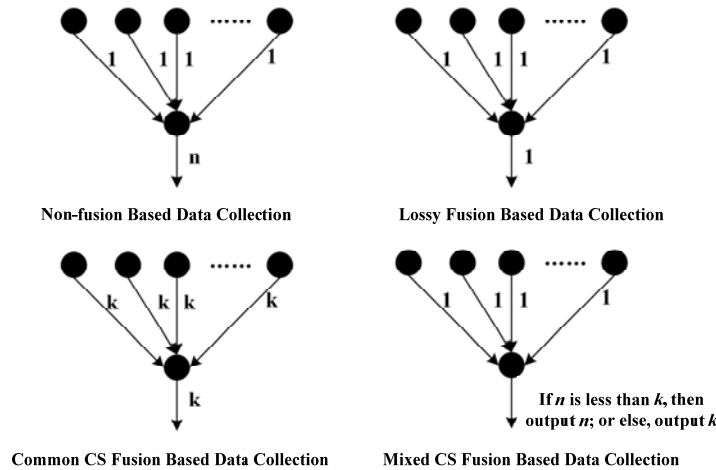


Fig.1. Data Collection under Different Fusion Modes

### Experiment Result Comparison

The energy consumption and the recovery error of the scheme proposed in this article under the condition of different node quantities are as shown in Figure 3. Obviously, the energy consumption is inversely proportional to the recovery error, namely: the energy consumption is increased along with the reduction of the recovery error, vice versa. The energy consumption increase means that the communication overhead of the node is increased and also indicates that more forwarding nodes

are included in the mixed CS fusion mechanism and the volume of the transmitted data is accordingly increased, in order words, the data recovery quality is ensured at the cost of more energy consumption. Additionally, under the condition of the same recovery error, when the node quantity is 300, the energy consumption is significantly reduced, because the scheme proposed in this article can be used to more rapidly and effectively divide the nodes in the network when there are relatively less nodes, and the algorithm also has less iteration processes and fast convergence speed which are favorable for saving energy.

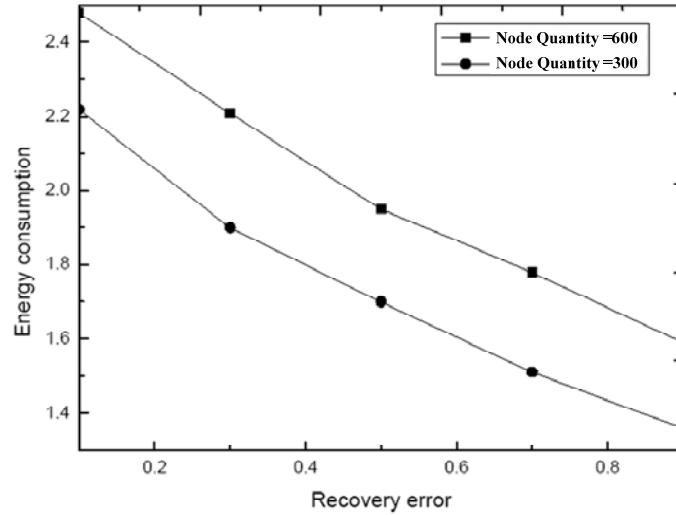


Fig.3. Energy Consumption VS Recovery Error

The comparison of the energy consumption for the data collection achieved by common CS fusion and mixed CS fusion under different node quantities is as shown in Figure 4. Obviously, along with the increase of the fusion factor  $k$ , the energy consumptions under two fusion modes are both increased, but the energy consumption under mixed CS fusion is always less than that under common CS fusion, because the increase of  $k$  indicates that more nodes in the network are converted into fusion nodes to adopt CS encoding for data transmission in order to reduce the volume of the transmitted data and save energy.

Additionally, at the same  $k$ , the energy consumption for the data fusion when the node quantity is 600 is significantly more than that when the node quantity is 300, because the probability for an optional node to become a fusion node when the node quantity is 300 is greatly higher than that when the node quantity is 600, namely: along with the reduction of the node quantity, more nodes may adopt CS encoding for data transmission to ensure energy effectiveness.

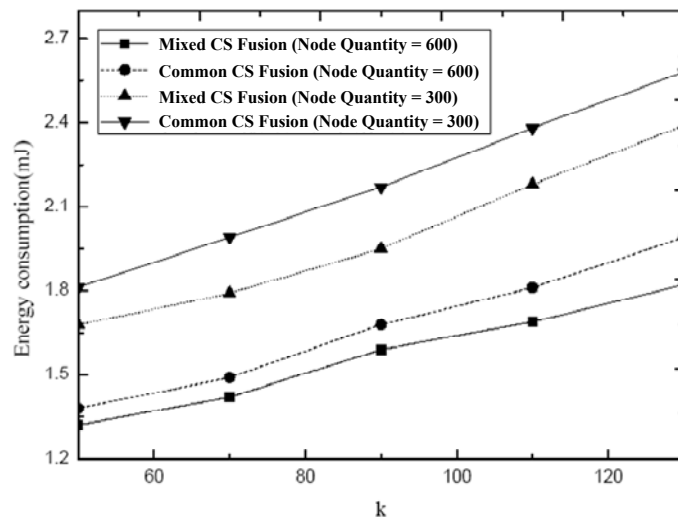


Fig.4. Comparison of Energy Consumptions under Different Fusion Modes

## Conclusion and Future Work

Data fusion is a basic research problem in the field of wireless sensor network. A compressible data fusion scheme with minimized energy consumption is proposed in this article on the basis of the compressed sensing theory. Specifically, the influence of route and mixed CS fusion on energy optimization is considered in this scheme, and a data fusion tree with optimized energy is obtained through network partitioning method. The result of the simulation experiment shows that the method proposed in this article is superior to the traditional methods in the aspect of data fusion accuracy and energy effectiveness. In future, we will lay the research emphasis on the following aspects: 1) research safety and reliability problems in data fusion on the basis of the compressed sensing theory; 2) in allusion to data loss, adopt the compressed sensing technology to model the data fusion problem into sparse signal recovery problem.

## References

- [1] Lv Z, Halawani A, Fen S, et al. Touch-less Interactive Augmented Reality Game on Vision Based Wearable Device[J]. *Personal and Ubiquitous Computing*, 2015, 19(3): 551-567.
- [2] Gu W, Lv Z, Hao M. Change detection method for remote sensing images based on an improved Markov random field[J]. *Multimedia Tools and Applications*, 2015: 1-16.
- [3] Chen Z, Huang W, Lv Z. Towards a face recognition method based on uncorrelated discriminant sparse preserving projection[J]. *Multimedia Tools and Applications*, 2015: 1-15.
- [4] Jiang D, Ying X, Han Y, et al. Collaborative multi-hop routing in cognitive wireless networks[J]. *Wireless Personal Communications*, 2015: 1-23.