

文章编号:1001-9081(2010)03-0729-04

# 无线传感器网络中基于区域相关性的自组织成簇算法

李 玮<sup>1,2</sup>, 胡玉鹏<sup>1</sup>

(1. 湖南大学 软件学院, 长沙 410082; 2. 湖南大学 计算机与通信学院, 长沙 410082)

(rj\_wli@hnu.cn)

**摘 要:**无线传感器网络资源有限,信息量大,通常采用分簇压缩减少传输量。针对传感器网络中的小波压缩,提出了一种基于相关区域自组织的成簇算法。该算法利用实际区域数据的相关性进行分簇,在簇头进行小波数据压缩的同时进行相关性检测,动态调整簇结构,保证簇内节点的相关性较好;同时在 Sink 分析簇间节点数据相关性,形成相关性好的大规模簇,进一步提高较长时间内的压缩效率。理论分析和实验仿真表明,该算法能尽可能地利用节点数据的时间和空间相关性去除冗余数据,提高小波数据压缩效率,降低了网络的能耗。

**关键词:**无线传感器网络;相关性;成簇;小波压缩

**中图分类号:** TP393 **文献标志码:** A

## Self-organized clustering algorithm based on regional relevance in wireless sensor network

LI Wei<sup>1,2</sup>, HU Yu-peng<sup>1</sup>

(1. Software School, Hunan University, Changsha Hunan 410082, China;

2. School of Computer and Communication, Hunan University, Changsha Hunan 410082, China)

**Abstract:** In view of limited resources and amounts of data in Wireless Sensor Network (WSN), clustering compressions are always employed to reduce transmission packets. Based on the regional relevance, a self-organized clustering algorithm was proposed for the wavelet compression in sensor network. The algorithm used the correlation of actual regional data for clustering. Actually, during the wavelet data compression procedure, the data correlation was detected in the cluster head, so that to insure better correlation and adjust the clusters structure. Meanwhile, data dependence among the nodes was analyzed in Sink, to form large-scale clusters with better relevance, so that to improve the efficiency of wavelet compression in a long period. Theoretical and experimental results show that the proposed algorithm can eliminate the data redundancy by using temporal and spatial correlation as much as possible, improve the efficiency of wavelet data compression, and reduce network energy consumption.

**Key words:** Wireless Sensor Network (WSN); correlation; clustering; wavelet compression

## 0 引言

无线传感器网络(Wireless Sensor Network, WSN)在军事和民用领域有着广泛的应用,如战场监视、环境和交通监测、灾难救助等,是目前非常活跃的研究领域<sup>[1-2]</sup>之一。无线传感器网络具有节点数据庞大、能量、存储和带宽都非常有限等特点,因此降低能耗以延长网络生命周期成为传感器网络中的核心问题。随着集成电路工艺的进步,处理器和传感器模块的功耗变得很低,绝大部分能耗消耗在无线通信模块上,传输1比特数据所消耗的能量大约相当于执行1000条CPU指令<sup>[3]</sup>。因此,通过节点对数据的压缩处理来减少数据的传输量,能有效地降低网络能耗。

部署在自然环境中的传感器节点所采集的数据是存在一定相关性的。利用节点的相关性进行数据压缩能大大降低数据传输量,节省网络能耗以延长网络生命周期。小波是一种可同时表征信号时域和频域行为的数学工具<sup>[4]</sup>,具有多分辨率的特性,在不同的尺度或者说压缩比下,仍然能够保持信号的统计特性。进行小波变换压缩时,数据的相关程度越高,其小波变换后的能量越集中,越能提高高频小波系数的编码

效率,提高压缩比。然而以往的压缩方法对于数据的实时动态性没有进行考虑<sup>[5]</sup>,本文提出一种基于相关区域自组织的成簇算法,针对实际应用中节点数据的实时相关性,自组织动态成簇,尽可能地挖掘出相邻簇之间的相关性,形成大规模簇,提高较长时间内的平均数据压缩效率;并采用OMNeT++平台对提出的算法进行仿真实验。理论分析和模拟实验表明,该算法能够减少传输的数据量,降低网络能耗。

## 1 相关工作及预备知识

传统的成簇协议如LEACH<sup>[6]</sup>(Low-Energy Adaptive Clustering Hierarchy)是一种自适应的成簇拓扑算法,通过动态等概率选举簇头来平均网络节点能耗。LEACH-C<sup>[7]</sup>是对LEACH的改进, Sink根据节点位置及能量等信息计算出成簇方案。GAF<sup>[8]</sup>(Geographical Adaptive Fidelity)的虚拟单元格思想为成簇机制提供了新思路。但是这些成簇是比较粗略和不准确的,考虑的只是理想模型,没有考虑实际监测数据的相关性。且在实际应用中由于地形、障碍物、节点分布等原因,节点的相关性往往并不是按距离远近呈圆形分布的,而是呈一定的带状或其他形状分布,在距离较远的区域之间也可能

收稿日期:2009-09-11;修回日期:2009-11-03。

**作者简介:**李玮(1972-),男,湖南临澧人,讲师,博士研究生,主要研究方向:计算机网络、生物计算; 胡玉鹏(1981-),男,湖南衡阳人,讲师,博士,主要研究方向:传感器网络、网络数据挖掘。

存在很强的相关性,且节点相关性会随时间发生变化。如在环境监测等需长期监测,数据量大的应用中,如果能够把数据相关的节点分在一个簇里,则能更好地对数据进行压缩处理。

传感器网络中基于小波的数据压缩研究目前已有一些基础性的工作。文献[9]设计了一个用于结构化监测的WISDEN系统,WISDEN先在单个传感器节点内对数据进行小波压缩,然后将其传送到基站进行集中式的处理以减少通信开销与网络延时。文献[10-11]提出了一个称为DIMENSIONS的层次系统,先在底层的各个传感器节点对监测到的数据进行小波压缩,然后基于一个称为WavRoute的路由协议,由中间层的汇聚节点收集底层节点传来的数据,并在汇聚节点进行进一步的小波压缩后传送到上一层节点。文献[12]提出了基于5/3小波提升方案,通过在邻近的节点间交换信息,在数据传送到汇聚节点前分布式挖掘网络中数据的空间相关性,减少了冗余数据的传输。但是这些模型系统设计都没有考虑实际数据的相关性,使得小波压缩达不到理想效果。

在本文中,提出了基于相关区域的自组织成簇算法,为了方便讨论,简化模型如下。

1) Sink(基站)位置固定,所有节点同构,并且具有相同的通信半径。

2) 节点固定或移动性很低,网络拓扑结构比较稳定。

3) 信道自由竞争无冲突。

4) 每个节点维护簇头ID(H),簇头维护一个簇内节点表(CH)。

5) 相关性传递。节点A与节点B相关,节点B与节点C相关,则节点A与节点C也相关;节点A与节点B相关,节点B与节点C不相关,则节点A与节点C也不相关。

## 2 基于相关区域自组织的成簇算法

### 2.1 地理成簇

将 $N$ 个节点部署在监测区域后,各节点通过相关定位技术(如GPS等)获得自身位置信息。Sink随机确定 $K$ 个节点为簇头: $H_{h_1}, H_{h_2}, \dots, H_{h_K}$ ,其中 $1 \leq h_{i=1,2,\dots,K} \leq N$ 且互不相同。假设节点 $r$ 的位置坐标为 $(x_r, y_r)$ ,则节点 $r$ 归属到 $H_i$ 簇头所在簇:当且仅当 $(x_r - x_{H_i})^2 + (y_r - y_{H_i})^2 = \min_{1 \leq j \leq K} \{(x_r - x_{H_j})^2 + (y_r - y_{H_j})^2\}$ 。这些计算均在Sink中完成,此后将这些成簇信息进行广播。节点收到信息后进行对比,确定自己是否为簇头以及所属簇ID,维护邻居节点表和簇内节点表。

### 2.2 算法描述

地理成簇只是简单地将最近的节点划分到同一个簇,并不能保证节点实际数据的相关性,因此,为了提高小波数据压缩的效果,有必要根据实时数据相关性对簇结构进行调整,保证簇内节点相关性较好。

首先,给出几个算法中要用到的定义。

普通节点 包括簇头和簇内节点,用集合 $G$ 表示。

孤立节点 除了普通节点以外的其他节点,用集合 $I$ 表示。

相关度 表示两个节点间数据的相关程度;设 $[x_{1,1}, x_{1,2}, \dots, x_{1,n}]$ 和 $[x_{2,1}, x_{2,2}, \dots, x_{2,n}]$ 分别为节点 $X_1$ 和节点 $X_2$ 的具有 $n$ 个数据的序列,则节点 $X_1$ 和 $X_2$ 的相关度为:

$$\text{Corr}(X_1, X_2) = \frac{L_{12}(X_1, X_2)}{\sqrt{L_1(X_1) \times L_2(X_2)}};$$

$$L_1(X_1) \neq 0, L_2(X_2) \neq 0$$

其中:

$$L_{12}(X_1, X_2) = \sum_{i=1}^n (x_{1,i} - \bar{X}_1)(x_{2,i} - \bar{X}_2)$$

$$L_1(X_1) = \sum_{i=1}^n (x_{1,i} - \bar{X}_1)^2$$

$$L_2(X_2) = \sum_{i=1}^n (x_{2,i} - \bar{X}_2)^2$$

$$\bar{X}_1 = \frac{1}{n} \sum_{i=1}^n x_{1,i}$$

$$\bar{X}_2 = \frac{1}{n} \sum_{i=1}^n x_{2,i}$$

两节点相关 两个节点的相关度大于某个阈值(可以由用户或专家来设置)。对于给定的阈值 $\rho$ ,如果 $|\text{Corr}(X_1, X_2)| \geq \rho$ ,则称节点 $X_1$ 和 $X_2$ 是相关的,否则不相关。在进行简单地地理成簇后,每个节点都成为普通节点,维护自己的邻居节点表、簇内节点表及簇头ID。

基于相关区域自组织成簇算法主要是根据实时数据的相关性动态对簇结构进行及时调整,提高簇内节点数据的相关性和小波压缩效率。算法主要分为簇内调整和簇间调整,描述如下。

#### 1) 簇内调整。

地理成簇;

for 任意节点 $N_i$

if  $N_i \in G$

// $N_i$ 为普通节点

$N_i$ 发送数据给簇头 $H_j$ ;

$H_j$ 计算两者相关性 $\text{Corr}(N_i, H_j)$ ;

if  $\text{Corr}(N_i, H_j) < \rho$

if  $H_j$ 的簇头节点表中 $\text{CH}(N_i) > 1$

$\text{CH}(N_i) = 1$ ;

else

$H_j$ 从 $\text{CH}$ 中删除 $N_i$ ,并发送 $N_i$ 的退簇报文 $ET_i$ ;

else

$\text{CH}(N_i)++$ ;

if  $N_i$ 收到自己的退簇报文 $ET_i$

$N_i$ 清空 $H_i$ 成为孤立节点;

else if  $N_i \in I$

// $N_i$ 为孤立节点

$N_i$ 对邻居节点广播数据;

for  $N_i$ 任意的邻居节点 $N_p$

if  $N_p \in G$

// $N_p$ 为普通节点

$N_p$ 计算两者相关性 $\text{Corr}(N_i, N_p)$ ;

if  $\text{Corr}(N_i, N_p) > \rho$

$N_p$ 给 $N_i$ 发送邀请加入簇报文 $CT_i$ ;

if  $N_i$ 收到 $N_p$ 邀请加入簇报文 $CT_i$

$N_i$ 给 $H_p$ 发送加入报文 $AT_i$ ,成为普通节点;

// $H_p$ 为 $N_p$ 的簇头

$H_p$ 收到 $N_i$ 的加入报文 $AT_i$ ,更新 $\text{CH}$ ;

if  $N_i$ 未收到任何邀请加入报文

$N_i$ 以一定概率 $P$ 升级为簇头,成为普通节点;

至此,算法簇内调整基本完成。但是如果两相邻簇之间存在相关性,则可以利用这种相关性提高压缩效率,为此将两个簇合并生成一个新的簇,簇结构更大,更加有利于进行小波压缩。

#### 2) 簇间调整。

Sink接收数据并还原;

Sink计算两个相邻簇簇头数据相关性 $\text{Corr}(H_i, H_j)$ ;

if  $\text{Corr}(H_i, H_j) > \rho$

给簇头 $H_i, H_j$ 发送合并簇报文;

$H_i$ 和 $H_j$ 收到合并簇报文;

if  $H_i$  和  $H_j$  簇内节点数小于最大负载节点数  
重新选举簇头  $H_k$  并更新  $CH$ ;

由于簇头节点能量负载比簇内节点重,如果簇头节点固定,簇头会因为能量的过度消耗而快速死亡,导致簇内节点无法与 Sink 正常通信。因此,每隔一定周期  $T$ ,簇内就进行簇头节点的重新选举。在一个簇内的簇头重新选举时,依据每个节点的能量及邻居节点的相关个数重新选择簇头。

另外,在退簇的时候,算法要求进行两次数据相关性计算是为了避免因异常数据而导致不相关引起不必要的簇结构调整,导致能量浪费。

### 3 性能评价及实验仿真分析

#### 3.1 性能分析

基于相关区域自组织成簇算法是针对在环境监测等需长期监测的应用提出的一种成簇算法,通过小波数据压缩减少传输数据来降低能耗。在 2.1 节中地理成簇同 LEACH 类似,但是只是简单地将距离较小的节点归为一簇,不能保证簇内节点的相关性。因此,在算法中利用实时节点数据之间的相关性来动态地调整簇结构,使得簇内节点具有较高的相关性,提高簇头对数据的压缩效率。同时,简单的地理成簇和动态调整局限在小范围内的相关性节点,而簇间结构调整则使得在较大的范围内的相关节点的能在同一个簇内,以得到更高的压缩效率,对于大规模无线传感器网络有着重要的意义。而小波变换的特点是:传感数据经变换后,绝大部分能量集中在低频系数上,高频部分的大量系数的值趋于 0,通过对高频部分系数进行量化处理和二元编码来提高压缩效率和压缩算法的鲁棒性。其中低频系数代表的是节点数据能量大小,高频系数代表的是数据变化细节,因此,节点数据越相关,高频系数就越趋于 0,所取得的压缩效果就越好。如图 1 所示,相关区域自组织成簇算法在进行动态调整簇结构时将节点 E 从簇 A 换到了簇 B,簇 C 和簇 D 合并成簇 F。

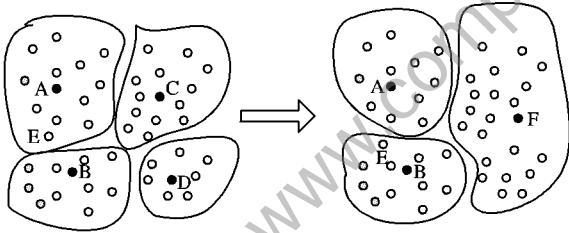


图1 相关区域自组织成簇示意图

下面分析该算法簇内调整对能耗的影响:首先,假设所发送的数据包同样大小,均为  $L$  bit,这样不影响模型整体分析的合理性。根据传感器节点的收发器能量模型,一个  $L$  bit 的分组传输距离为  $d$ ,则发送能耗为  $E_{Tx} = E_{elec} \times L + \varepsilon \times L \times d^2$ ,接收能耗为  $E_{Rx} = E_{elec} \times L$ ,其中,  $E_{elec}$  表示发射和接收电路每发送和接收单位 bit 的能耗,单位为 J/bit;  $\varepsilon$  表示放大器将每 bit 传输单位平方米所耗的能量,单位为 J/(bit · m<sup>2</sup>)。设节点在簇内调整前的压缩比为  $m$ ,簇内调整后的压缩比为  $m_c$ ,区域面积大小为  $S$ ,节点总数为  $N$ ,簇头个数为  $K$ ,节点在簇内调整后稳定阶段发包数为  $M$ ,簇头到 Sink 的跳数为  $D_{toSink}$ 。因为调整前后簇内节点能耗不变,因此只考虑簇头能耗的变化。未进行簇内调整簇头能耗为:

$$E = \frac{NM}{Km} (E_{elec} \times L + \varepsilon \times L \times d^2) \times D_{toSink}$$

进行簇内调整簇头能耗为:

$$E_c = \frac{NM}{Km_c} (E_{elec} \times L + \varepsilon \times L \times d^2) \times D_{toSink}$$

调整能耗为:  $E_{adjust} = E_{退簇} + E_{邀请} + E_{计算}$

其中:  $E_{退簇} = E_{Tx} + E_{Rx} = 2E_{elec} \times L + \varepsilon \times L \times d^2$ ,  $E_{邀请} = \frac{S}{\pi d^2} NE_{Tx} + 2(E_{Tx} + E_{Rx}) = \left( \frac{SN}{\pi d^2} + 2 \right) (E_{elec} \times L + \varepsilon \times L \times d^2) + 2E_{elec} \times L$ 。

因为计算能耗相比发射能耗小得多,所以,簇头节约能耗为:

$$E_{save} = E - (E_c + E_{退簇} + E_{邀请}) = \left( \frac{N}{K} \times \left( \frac{M}{m} - \frac{M}{m_c} \right) \times \right.$$

$$\left. D_{toSink} - \frac{\pi d^2 N}{S} - 3 \right) E_{Tx} - 3E_{Rx} \quad (1)$$

簇间调整对能耗的影响,设簇间调整后的压缩比为  $m_c'$ ,簇头节约能耗为:

$$E_{save}' = \left( 2 \frac{N}{K} \times \left( \frac{M}{m} - \frac{M}{m_c'} \right) \times D_{toSink} \right) - (E_{Tx} + E_{Rx}) \times$$

$$D_{toSink} - E_{重选簇头} \quad (2)$$

由式(1)、(2)可看出,随着网络密度的增加和网络规模的扩大,簇头离 Sink 的距离越远,跳数越多,节约能耗也越多。同时,节点数据间相关性越稳定,即  $M$  值越大,节约能耗也会增加。

值得一提的是,相关区域自组织成簇方法不仅在压缩方面表现较好,而且在融合、分析、理解、决策和查询等方面也表现出较大的优势。

#### 3.2 实验仿真

实验数据使用 Intel-伯克利大学联合研究实验室利用传感器节点采集的数据,测试平台如下:1) CPU 为 Intel P4 3.06 GHz 和内存为 2 GB 的计算机;2) 操作系统为 Windows XP;3) 利用面向对象的开源仿真软件 OMNeT++ 3.3 Win32 (exe) 和基于其上支持无线和移动仿真框架 mobility-fw2.0p3。

模拟实验假设传感器分布在一个  $300 \times 300$  的区域内,设  $N$  表示节点总数,  $\rho$  表示衡量相关度值。取  $N = 200$ ,  $\rho = 0.6$ ,对于不同的均方差 (Mean Square Error, MSE) 要求,实验结果如图 2、3 所示。

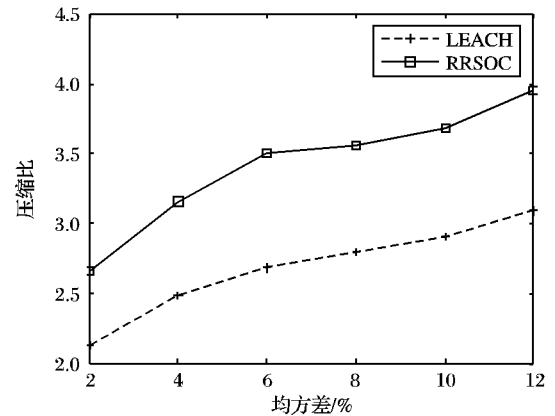


图2 不同均方差时的压缩比

图 2 表明:在相同 MSE 要求下,RRSOC (Regional Relevance based Self-Organized Clustering algorithm) 算法产生的平均压缩比 LEACH 的大,能更有效地压缩数据,减少传输数据量。同时图 4 表明:在相同 MSE 要求下,RRSOC 算法比 LEACH 算法更能节省能耗。这是因为 RRSOC 算法能够将相关性强的数据节点分在同一个簇,有利于提高小波变换编码效率,减少传输数据和节省能耗。

图 4 表明:随着节点总数的增加,RRSOC 算法和 LEACH

算法的压缩比都增加,这是因为节点总数增加,节点数据的冗余量也增大。但是,RRSOC 算法比 LEACH 算法的压缩比增加更快,说明 RRSOC 能更有效地减少数据冗余。图5表明:随着节点总数的增加,RRSOC 算法的能耗始终比 LEACH 算法的能耗低,且两者差值呈增大趋势。但是由于 RRSOC 动态调整簇结构需消耗一定能量,所以跟压缩比变化相比不是那么明显。图6表明:随着节点总数的增加,RRSOC 算法的数据精度始终比 LEACH 算法的精度高,这是因为 RRSOC 算法保证了簇内数据的相关性较好,在进行小波变换后量化编码时效率较高,所以数据精度较高。

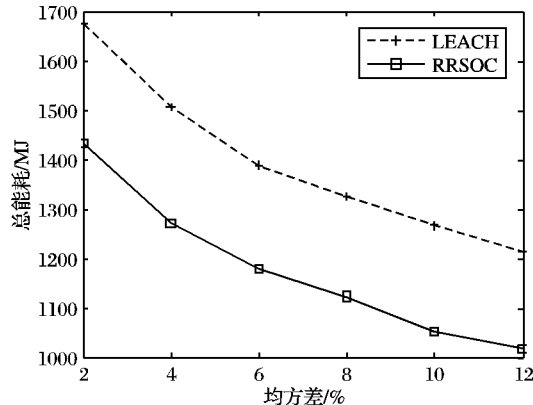


图3 不同均方差时的能耗

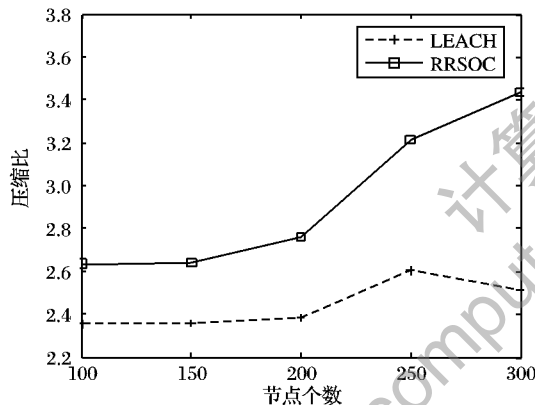


图4 不同节点数时的压缩比

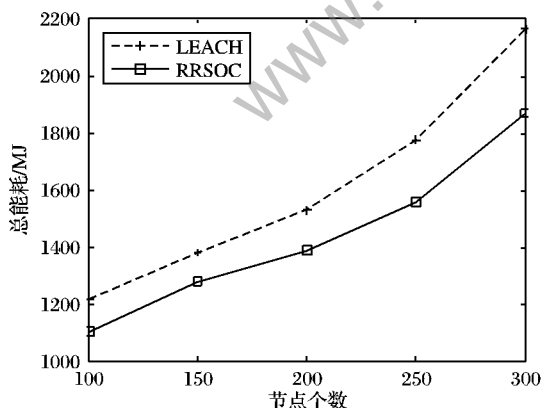


图5 不同节点数时的能耗

#### 4 结语

本文针对无线传感器节点实际数据相关性提出一种基于相关区域自组织成簇算法,该算法在网络稳定的情况下能显著提高小波数据压缩效率,且能对于网络变化作出及时动态簇结构调整,保证簇内节点具有较高的相关性和小波压缩效率。理论分析和实验仿真也验证了该算法的性能。

对节点的剩余能量、相关度等因素对簇头选举及网络性能的影响进行量化分析,将是下一步研究的重点。

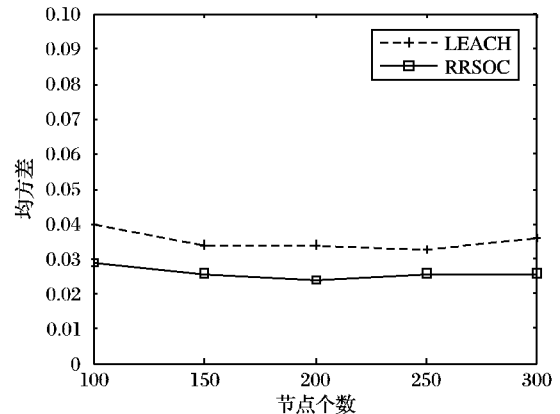


图6 不同节点数时的均方差

#### 参考文献:

- [1] LINDSEY S, RAGHAVENDRA C, SIVALINGAM K M. Data gathering algorithms in sensor networks using energy metrics [J]. IEEE Transactions on Parallel and Distributed Systems, 2002, 13(9): 924-935.
- [2] 李建中, 李金宝, 石胜飞. 传感器网络及其数据管理的概念、问题与进展[J]. 软件学报, 2003, 14(10): 1717-1727.
- [3] CHEN M O, FOWLER M L. Data compression trade-offs in sensor networks [C]// Proceedings of the SPIE Conference on Mathematics of Data/Image Coding, Compression, and Encryption. Denver, CO: SPIE, 2004, 5561: 96-107.
- [4] 成礼智, 王红霞, 罗永. 小波的理论与应用[M]. 北京: 科学出版社, 2004.
- [5] 周四望, 林亚平, 张建明等. 传感器网络中基于环模型的小波数据压缩算法[J]. 软件学报, 2007, 18(3): 669-680.
- [6] HEINZELMAN W, CHANDRAKASAN A, BALAKRISHNAN H. Energy-efficient communication protocol for wireless microsensor networks [C]// Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. Washington, DC: IEEE Computer Society, 2000: 3005-3014.
- [7] HEINZELMAN W B, CHANDRAKASAN A P, BALAKRISHNAN H. An application-specific protocol architecture for wireless microsensor networks [J]. IEEE Transactions on Wireless Communications, 2002, 1(4): 660-670.
- [8] XU Y, HEIDEMANN J, ESTRIN D. Geography-informed energy conservation for Ad Hoc routing [C]// Proceedings of the 7th Annual International Conference on Mobile Computing and Networking. Rome, Italy: [s. n.], 2001: 70-84.
- [9] XU N, RANGWALA S, CHINTALAPUDI K K, et al. A wireless sensor network for structural monitoring [C]// Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2004: 13-24.
- [10] GANESAN D, ESTRIN D, HEIDEMANN J. Dimensions: Why do we need a new data handling architecture for sensor networks? [J]. SIGCOMM Computer Communication Review, 2003, 33(1): 143-148.
- [11] GANESAN D, GREENSTEIN D, PERELYUBSKIY D, et al. An evaluation of multi-resolution storage for sensor networks [C]// Proceedings of the 1st International Conference on Embedded Networked Sensor Systems. New York: ACM Press, 2003: 89-102.
- [12] CIANCIO A, ORTEGA A. A distributed wavelet compression algorithm for wireless sensor networks using lifting [C]// Proceedings of the 2004 International Conference on Acoustics, Speech, and Signal Processing. Piscataway: IEEE Computer Society, 2004: 633-636.