

doi: 10.11823/j.issn.1674-5795.2019.03.05

无线传感网络中大数据量压缩感知编解码算法

郝有田¹, 宋萍¹, 郝创博^{1,2}

(1. 北京理工大学, 北京 100081; 2. 北京京航计算机通讯研究所, 北京 100074)

摘要: 针对无线传感网络传输数据的冗余性问题, 提出了一种逐级压缩感知编解码算法。该压缩算法将原始信号按照相同或不同字典的稀疏性进行信号逐级分解, 并利用伯努利观察矩阵对分解后的信号进行压缩编码, 在压缩过程中生成字典掩模。压缩后的信号和掩模回传至终端。终端根据编码信息、稀疏字典、字典掩模对数据进行逐级恢复。该数据压缩方法较传统的编码压缩技术具有更强的鲁棒性, 对丢包不敏感; 较原压缩感知算法, 节省无线网络的传输带宽, 提高数据信息获取的实时性。

关键词: 无线传感器网络; 压缩感知; 编解码; 恢复算法

中图分类号: TB97

文献标识码: A

文章编号: 1674-5795(2019)03-0028-06

Compressed Sensing Codec Algorithm for Big Data in Wireless Sensor Networks

QIE Youtian¹, SONG Ping¹, HAO Chuangbo^{1,2}

(1. Beijing Institute of Technology, Beijing 100081, China;

2. Beijing Jinghang Computer Communication Research Institute, Beijing 100074, China)

Abstract: Aiming at the redundancy problem in wireless sensor network, a step-by-step compression-aware codec algorithm is proposed. The algorithm decomposes the original signal according to the sparsity of the same or different dictionaries, and compresses and encodes the decomposed signal by using the Bernoulli observation matrix to generate a dictionary mask during the compression process. The compressed signal and mask are sent back to the terminal. The terminal restores the data step by step according to the coding information, the sparse dictionary and the dictionary mask. This data compression method is more robust than traditional coding compression technology and is not sensitive to packet loss. Compared with the original compressed sensing algorithm, it saves the transmission of wireless network bandwidth and improves the real-time performance of data information acquisition.

Key words: wireless sensor networks; compressed sensing; codec; recovery algorithm

0 引言

无线传感器测试网络采集的数据大部分为温度、湿度、振动、位移、声音等自然界中的一维数据。无线传感器网络中的数据压缩编码主要针对这类一维信号开展。压缩编码之所以可以做到数据压缩, 实质上利用了无线传感器网络的数据相关性。这种相关性包括时间相关性和空间相关性。因此, 相应的压缩算法也可以分为基于时间相关性的数据压缩算法、基于空间相关性的数据压缩算法, 以及基于时空相关性的数据压缩算法。

这些典型的数据压缩算法和设计思路多数仅考虑了数据的冗余性, 没有考虑无线传输中的不稳定性和丢包现象。为此, Luo 和 Lee 等人将压缩感知算法^[1-2]引入无线传感器网络领域, 在成功剔除了数据空间冗余的同时保证了数据对丢包的不敏感性。Xiang L 等人

结合小波变换, 通过压缩感知算法剔除了数据的时空冗余^[3]。这些算法利用了压缩感知算法观测矩阵设计的灵活性, 提高了信号重构的鲁棒性, 但是往往需要较高维度的观测信号保证数据的重构精度, 影响了数据的压缩率和终端显示的实时性。

针对以上问题, 本文以时间冗余剔除问题为例, 根据信号稀疏域的系数对信号进行分级压缩感知, 利用字典掩模方法降低了重构信号所需的观测信号维度, 同时利用分级重构和多稀疏域重构算法, 提高了数据重构的实时性和准确度, 节约了网络的传输带宽。

1 压缩感知中信号分解可行性分析

在真实环境中, 采集的自然界物理量多数不是严格稀疏的。例如振动信号在幅值谱和功率谱中, 如图 1 所示, 大部分频率上的信号是 0 或者近似于 0, 仅有少

数几个频率上的值较大。对此类信号采用压缩感知进行数据冗余剔除时,也是一个滤波过程,将在主稀疏基中次要的信号滤去,因此压缩感知在实际应用中多为一种有损压缩过程。

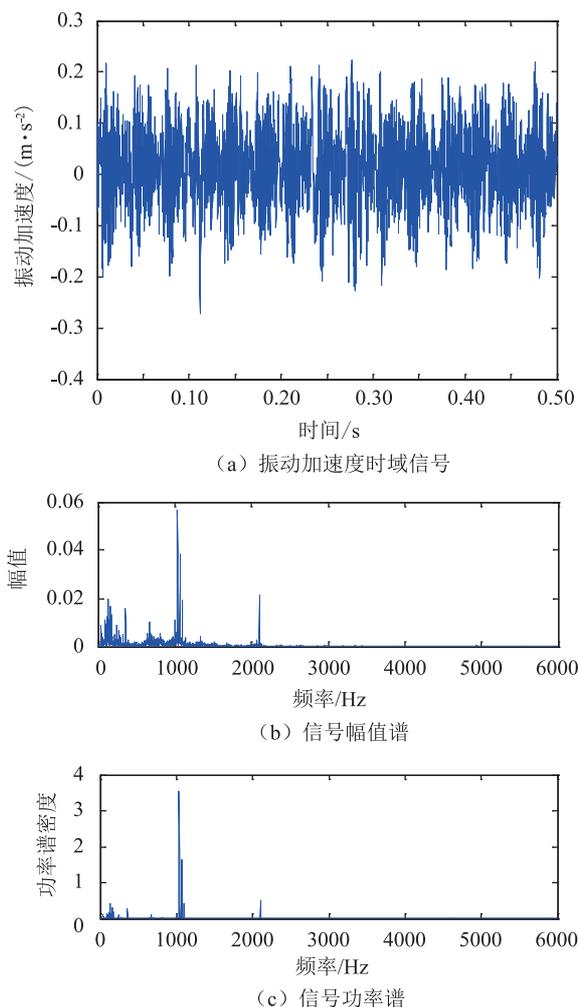


图1 振动加速度时域信号、频谱和功率谱

对于 Orthogonal Matching Pursuit (简称 OMP 算法) 这种常用的信号还原方法,需要测量矩阵满足 $M \geq cK_{sp} \log(N/K_{sp})$ 才能较好的还原原始信号。由于测量矩阵行数 M 是降维后信号的长度,可以利用式(1)计算压缩感知的最大压缩比。

$$R_{\text{comp}} = M/N \geq cK_{sp} \log(N/K_{sp})/N \quad (1)$$

式中: N 为原信号长度; K_{sp} 为稀疏度; M 为降维后观测矩阵的秩。由于 c 为一个很小的常数,且 $1 \leq K_{sp} \ll N$, 将式(1)右端进行近似化简。

$$\begin{aligned} cK \log(N/K_{sp})/N &= [cK_{sp} \log N - cK_{sp} \log K_{sp}]/N \\ &= (c \log N - c \log K_{sp})K_{sp}/N \\ &\approx c \log N \cdot K_{sp}/N \end{aligned} \quad (2)$$

$$R_{\text{comp}} = M/N \geq (cK_{sp} \log N)/N \quad (3)$$

可见,其压缩率与稀疏度近似成正比,与信号的稀疏度相关。因此,当原始信号长度 N 一定时,可以基于稀疏度将信号进行分解,将一个稀疏信号分解为几个正交稀疏信号相加的形式,然后分别对这些信号进行压缩编码,这样并不影响总信号的压缩率及压缩编码长度,如式(4),而分解后的信号可提高重构的实时性和精度,本文将具体设计逐级分解、压缩、重构方法,其总体编解码流程如图2所示。

$$\begin{aligned} K_{sp} &= K_{sp1} + K_{sp2} + \dots + K_{spn} \\ R_{\text{comp}} &= (M_1 + M_2 + \dots + M_n)/N \\ &\geq c(K_{sp1} + K_{sp2} + \dots + K_{spn})(\log N)/N \\ &= (cK_{sp} \log N)/N \end{aligned} \quad (4)$$

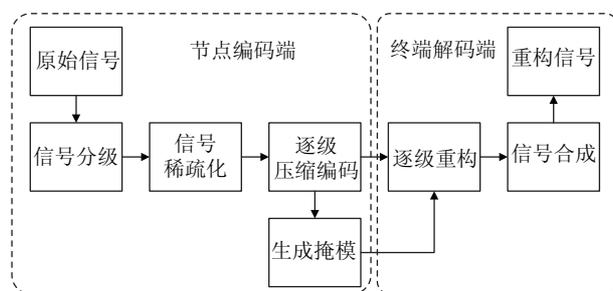


图2 总体编解码流程

2 基于稀疏性的信号逐级分解及恢复

2.1 基于稀疏向量系数的逐级分解与稀疏化方法

通过上述描述可知,可以在对总信号压缩率影响不大的前提下,以稀疏基下信号稀疏表示的分量系数大小为标准对信号进行分解。其基本思路为:首先从总信号中分解出最主要、最稀疏的信号,之后从残差中分解出较次要的信号,依次类推。一般的自然界物理量是近似稀疏的,因此,往往最开始分解出的信号是稀疏的,而随着分解的逐级进行,信号的稀疏性会越来越差,能量也会越来越小。剩余最后的残差信号被忽略。用公式表示为

$$\begin{aligned} X &= \Psi \alpha + r = \Psi \sum_i \alpha_i + \Psi \alpha_r \\ X_i &= \Psi \alpha_i \end{aligned} \quad (5)$$

式中: X 为原始信号; α 为原始信号在稀疏基 Ψ 下的稀疏表示; r 为剩余残差; α_r 为残差信号稀疏表示。由于稀疏基为正交稀疏基,为了各分信号间的信息相对独立,各分信号应满足式(6),分信号的稀疏基原子系数相互独立,任意两个分信号之间的相关性为0,即

$$\alpha_i(k) \neq 0 \Rightarrow \forall j, \alpha_j(k) = 0 \quad (6)$$

如此，每个分信号稀疏度的和为原始稀疏信号的稀疏度

$$\|\alpha\|_0 = \sum_i \|\alpha_i\|_0 \quad (7)$$

为了达到能量逐级递减的效果，分解时应保证每个分信号中的最大系数小于上一级中的最小系数。

$$\max(\alpha_i) < \min(\alpha_{i-1}) \quad (8)$$

这里设计采用自适应阈值法定义每一级 α_i 元素的取值范围

$$\sigma \cdot \max(\alpha_r) \leq \alpha_i(n) \leq \max(\alpha_r) \quad (9)$$

式中： σ 为该级分信号的稀疏向量系数下限占最大残差信号稀疏向量系数的比例。第 i 级残差信号稀疏表示可用式(10)计算。

$$\alpha_r = \alpha - \sum_{j=1}^{i-1} \alpha_j \quad (10)$$

信号的分解步骤流程图如图 3。①首先将原始信号进行稀疏转换，计算原始信号的稀疏表示 α ；②初始化残差信号的稀疏表示 $\alpha_r = \alpha$ ，分解级数为 $i = 1$ ；③计算 α_i 。对残差 α_r 中的每个元素进行遍历。若其中的元素 $\alpha_r(j)$ 大于 $\sigma \max(\alpha_r)$ ，则 $\alpha_i(j) = \alpha_r(j)$ ， $\alpha'_r(j) = 0$ ；否则， $\alpha_i(j) = 0$ ， $\alpha'_r(j) = \alpha_r(j)$ 。在遍历的过程中记录新的分信号的稀疏度 K_i ；④使用 α'_r 更新残差 α_r ，分解级数更新为 $i = i + 1$ ；⑤重复步骤 3 和步骤 4，直至分解级数达到某个阈值 n 或者残差信号的 p 范数小于某个阈值。

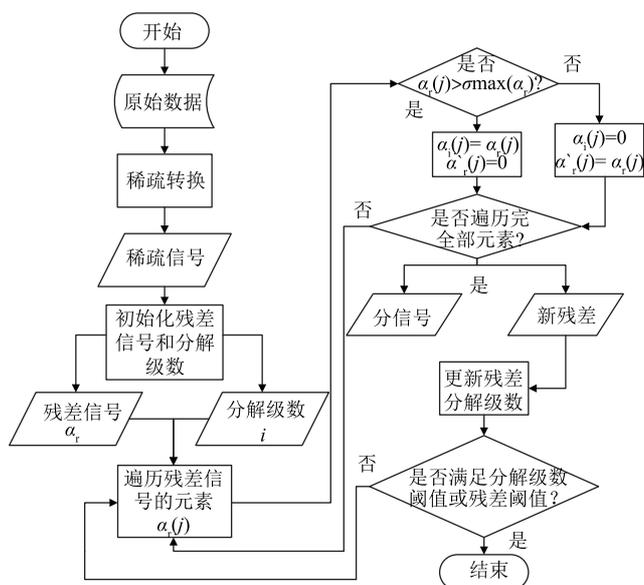


图 3 基于稀疏向量系数的分解流程

对信号完成分级后，原始信号被分解为各级的信号稀疏表示 $\alpha_1, \alpha_2, \alpha_3 \dots \alpha_n$ 。同样，可以从各级信号中按式(9)还原原始信号。

2.2 信号逐级编码压缩

1) 观测矩阵确定

各级信号的逐级压缩，同传统压缩感知的方法相同，需要首先确定观测矩阵。观测矩阵为一个与稀疏基不相关的测量矩阵与稀疏基的乘积。其中，信号的稀疏基是由信号本身属性决定。因此确定了测量矩阵，观测矩阵也相继确定。测量矩阵的大小为 $M \times N$ ， M 值决定了编码信号降维后的维度。 M 值越小，在完成压缩编码后，数据的大小越小，压缩率越低，压缩效果越好。但 M 值不能无限制的减少，一般来说，需要大于某一下限。若测量矩阵为高斯随机矩阵时，可以大概率还原原始信号^[4]。伯努利随机矩阵与高斯随机矩阵具有近似的普适性和重构能力，且较高斯随机矩阵来说更便于硬件计算和存储^[5]。因此在无线传感器网络中，选用伯努利随机矩阵作为压缩感知中的测量矩阵，如式(11)所示。伯努利随机矩阵的大小由所需压缩率与下限 $cK \log(N/K)$ 确定，其中， c 值根据信号的先验知识确定。

$$\varphi_{i,j} = \begin{cases} +1/\sqrt{M}, p = 0.5 \\ -1/\sqrt{M}, p = 0.5 \end{cases} \quad (11)$$

从上节信号分解情况可知，各级分信号的稀疏度是不相同的，往往级数越低的分信号越稀疏，测量矩阵所需的 M 值越小。若为每级分信号单独定义观测矩阵，容易造成硬件资源的浪费。为了不保存多余的观测矩阵，首先根据先验知识在网络节点和终端软件处保存相同的 M 值，足够大的总观测矩阵 $M_{\max} \times N$ ，且满足

$$M_{\max} > \max(M_1, M_2, \dots, M_n) \quad (12)$$

该观测矩阵是由一个足够大的伯努利随机矩阵与信号的先验稀疏基相乘所得。每一级分信号的观测矩阵作为该观测矩阵的子矩阵，其矩阵的列数与总观测矩阵相同，而行数则是根据各级分信号所需 M_i ，从总观测矩阵第一行起到第 M_i 行止，进行选取。如此，终端可以根据接收到分信号的降维信号维度，判断其观测矩阵，以便进行后续的数据重构。

2) 观测矩阵掩模设计

值得注意的是，由于信号的稀疏性，单对于某一分信号来说，由于大部分稀疏信号的元素为 0，在压缩

编码过程中真实参与计算的观测矩阵中的列向量很少。并且对于自然界的物理量信号，在一些常用域(如频域，小波域等)中稀疏变换后，其稀疏向量中的非0值所处的索引位置往往相近或相邻。为此，通过设计一种观测矩阵掩模向量，标记出编码时参与计算的观测矩阵中原子的大致区域或范围，并将该掩模向量一并传递至终端，降低重构时观测矩阵中参与投影的原子个数，减少重构时的计算量，提高编码压缩后信号还原的能力。

为了降低掩模向量的数据长度，每个掩模向量中的元素并不是表示单一观测矩阵原子参与编码的情况，而是对应于一个连续的观测矩阵原子集合。向量中元素取值为0或者1，0表示其对应的观测矩阵原子集合中没有参与计算的原子，1表示其对应的观测矩阵原子集合存在参与编码计算的原子，每个元素可以用1 bit表示。

2.3 基于掩模观测矩阵的 OMP 逐级信号重构算法

由于原始信号被分解为各级分信号分别进行压缩降维，在信号重构时，也同样采用逐级重构的方法。选用 OMP 算法为基础算法^[6]，为每一级分信号进行重构。OMP 算法作为压缩感知中信号重构经典算法，广泛应用于各种应用场景，具有良好的实时性，且实现简单。不同于传统的 OMP 算法，本节介绍的重构算法将考虑分级及掩模向量对残差的追踪过程的影响。

1) 基于掩模观测矩阵的 OMP 信号重构算法

本文采用基于掩模观测矩阵的 OMP 信号重构算法对各级分信号进行重构。与传统 OMP 不同，在计算残差与原子内积时，本算法需要在重构字典中剔除掩模信息未标记的原子和其他已重构分信号所用的原子。本算法的输入有观测矩阵 Θ 、采样向量 \mathbf{Y} 、掩模向量 Ω 、其他已重构分信号的稀疏向量 $[\alpha_1, \alpha_2, \dots, \alpha_{i-1}]$ 、内积终止阈值 Th ，输出信号为当前分信号 \mathbf{X}_i 的稀疏逼近 $\hat{\mathbf{X}}_i$ 。其具体算法步骤有：

步骤1：初始化：残差 $\mathbf{r}_0 = \mathbf{Y}$ ，索引集 A_0 为空，重建原子矩阵 \mathbf{S} 为空，迭代次数 t 为 0。

步骤2：查找可用原子集合中的原子与残差 \mathbf{r} 的最大内积值及对应的索引值，如式(13)。

$$|\langle \mathbf{r}, \boldsymbol{\theta}_{\lambda_t} \rangle| = \sup_{i \in E} |\langle \mathbf{y}, \boldsymbol{\theta}_i \rangle|$$

$$E = \{k | Y(\eta_k) = 1, \alpha_{j \in (1, 2, \dots, i-1)}(k) = 0, k \notin A\} \quad (13)$$

式中： $Y(\eta_k)$ 为在字典中索引为 k 的原子在掩模向量中对应的值； E 为可用原子索引的集合。

步骤3：判断最大内积值 $|\langle \mathbf{r}, \boldsymbol{\theta}_{\lambda_t} \rangle|$ 是否小于阈值

Th ，若满足则输出分信号 $\hat{\mathbf{X}}_i = \mathbf{S} \hat{\boldsymbol{\alpha}}_i$ ，停止迭代。

步骤4：更新索引集 $A_t = A_{t-1} \cup \{\lambda_t\}$ ，更新重建原子矩阵 $\mathbf{S}_t = [\mathbf{S}_{t-1}, \boldsymbol{\theta}_{\lambda_t}]$ 。

步骤5：计算最小二乘，得到稀疏向量，如式(14)。

$$\hat{\boldsymbol{\alpha}}_i = (\mathbf{S}_i^T \mathbf{S}_i)^{-1} \mathbf{S}_i^T \mathbf{y}_i \quad (14)$$

步骤6：更新残差 $\mathbf{r}_i = \mathbf{y} - \mathbf{S} \hat{\boldsymbol{\alpha}}_i$ ，更新迭代次数 $t = t + 1$ 。

重复执行步骤2与步骤6。2) 逐级重构信号

采用逐级重构信号的优势在于：单个分信号的编码数据传输完成后，便可以对原始信号进行重构并输出。随着重构完成的分信号越多，级数越高，原始信号的细节信息越多，失真越小。这种逐级重构的方式有别于传统的信号重构，不用接收完成所有压缩编码信息便可进行重构信号的输出。如此，重构信号输出过程将传统压缩感知从“无”到“有”的过程，转化为一个从“模糊”到“清晰”的过程。其逐级重构过程如图4所示。

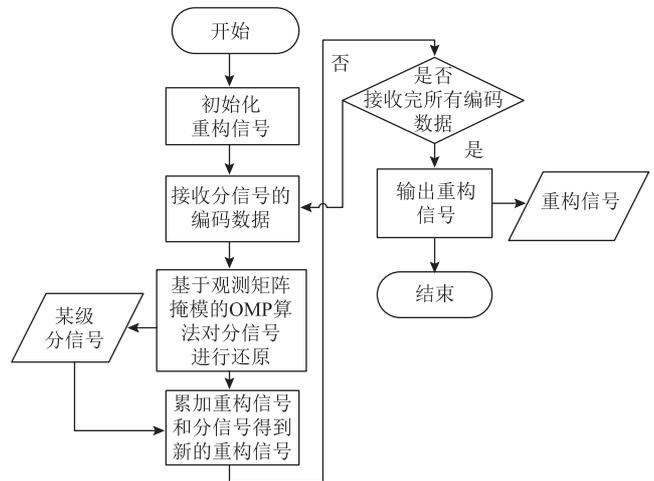


图4 分级重构信号流程

网络中首先传输级数低的分信号编码，因此首先重构出来的是稀疏向量系数最大的分信号，而后不断的累加级数较高的分信号，实现细节的修正，设完成第 j 级分信号的重构，则原始信号重构如式(15)所示。

$$\mathbf{X}_{\text{rec}} = \sum_{i=1}^j \mathbf{X}_i = \boldsymbol{\Psi} \sum_{i=1}^j \boldsymbol{\alpha}_i \quad (15)$$

3 试验验证

3.1 试验数据源选择

针对本文所设计的压缩感知算法，设计仿真试验

证明其有效性，并且在相同仿真参数和信号输入的环境下同传统的压缩感知算法进行对比。本节设计三组仿真试验，采用美国西储大学轴承数据集中随机抽取的部分轴承振动加速度数据作为原始信号，分别采用本文算法、不做稀疏处理的传统压缩感知算法和做稀疏处理的传统压缩感知算法进行不同压缩率下的信号重构试验，对比重构结果。

3.2 试验参数设置

数据采用美国西储大学轴承数据集中随机抽取部分轴承振动加速度数据如图 5。该数据的采样率为 12 ks/s，截取 6000 个连续采样点。运用本文所提压缩算法和传统的压缩算法分别在每个不同压缩率下进行 10 组试验对信号进行还原，还原后计算与原始信号的均方根误差。本文所提算法将原始信号分为 4 级傅里叶稀疏域的分信号，每一级按照式(9)进行划分， σ 取 0.5，信号重构采用上节中提到的具有掩模的分级重构方法。传统的压缩感知算法则设置相同的压缩率，并且也对源信号做相同的稀疏处理，采用傅里叶稀疏域和时域的综合字典进行信号 OMP 重构。

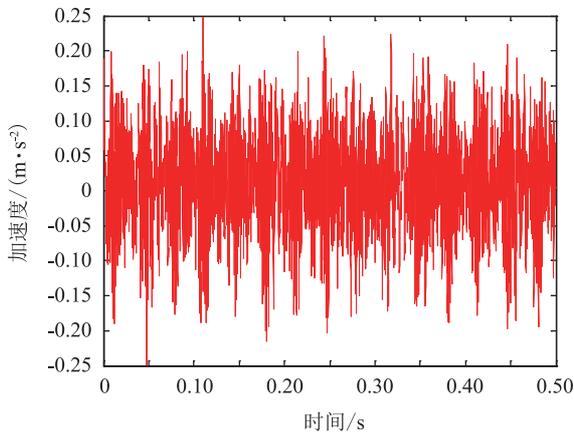


图 5 轴承振动加速度数据作为原始信号

3.3 实验结果

试验结果如图 6 所示。可以看到，本文所提到的算法在信号压缩及重构时相同压缩率的情况下具有一定优势。对比稀疏处理和不稀疏处理的算法，可见在压缩编码前若不滤去不稀疏的剩余残差信号就进行稀疏处理，其非稀疏量对信号追踪产生干扰，不利于信号重构。对比同样做了稀疏处理的传统压缩感知和本文算法，可见超低压缩率下，本文算法在重构后的误差散布较大，此时传统压缩感知的误差散布小但误差均值较大。随着编码数据量的增加，本文算法很快收

敛到一个较低的固定值，这个固定值是由于稀疏处理带来有损压缩的结果，是不可避免的。而传统压缩感知仍具有较高的误差，直至编码数据量大于一定程度才收敛到相同的固定值。

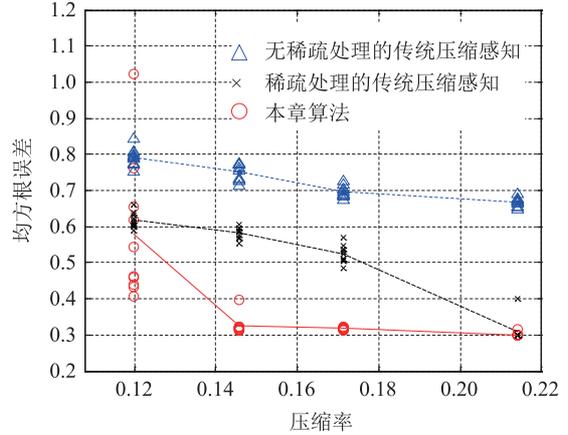


图 6 压缩率与重构误差的试验结果

4 结论

针对无线传感器网络所采集的数据存在数据大量冗余的问题，而传统的编码压缩算法又对丢包过于敏感，本文提出了一种基于压缩感知的信号分解压缩算法。通过分析压缩感知编码长度与稀疏性的关系证明了算法的可行性，设计算法的分级分解、压缩编码、分级重构等环节，开发出基于不同稀疏基的分级压缩感知算法。最后，通过对比试验验证了所提算法具有压缩感知编码优点的同时，较传统压缩感知算法，具有更高的压缩率、更好的实时性。

参 考 文 献

- [1] Luo C, Wu F, Sun J, et al. Compressive Data Gathering for Large-scale Wireless Sensor Networks [C]//International Conference on Mobile Computing and Networking, 2009: 145 - 156.
- [2] Lee S, Patten S, Sathiamoorthy M, et al. Spatially-Localized Compressed Sensing and Routing in Multi-hop Sensor Networks [C]// International Conference on Geosensor Networks, Springer-Verlag, Berlin, Heidelberg, 2009: 11 - 20.
- [3] Xiang L, Luo J, Rosenberg C. Compressed Data Aggregation: Energy-Efficient and High-Fidelity Data Collection [J]. IEEE/ACM Transactions on Networking, 2013, 21 (6): 1722 - 1735.
- [4] Candes E J, Tao T. Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies? [J]. IEEE

Transactions on Information Theory, 2006, 52 (12): 5406 - 5425.

- [5] Baraniuk R, Davenport M, Devore R, et al. A Simple Proof of the Restricted Isometry Property for Random Matrices[J]. Constructive Approximation, 2008, 28(3): 253 - 263.
- [6] Theis F J, Jung A, Püntonen C G, et al. Signal Recovery from Partial Information Via Orthogonal Matching Pursuit[J]. IEEE Trans. inform. theory, 2005, 15(2): 419 - 439.

收稿日期: 2018 - 11 - 19

基金项目: 国家“十三五”基础科研项目 (JCKY2016602B002, JCKY2016208B007); 国家“十三五”重大基础科研项目 (JCKY2017209A003)

作者简介

鄢有田(1994 -), 男, 硕士, 主要研究方向为无线传感器网络技术。



宋萍(1972 -), 女, 教授, 博士, 博士生导师, 主要从事感知与测控技术方面的研究。北京理工大学智能机器人与系统高精尖创新中心教授, 机械工程一级学科博士生导师, “仿生机器人与系统”教育部重点实验室副主任, 兼任国防科技工业科学技术委员会科技创新领域专业组成员、中国计算机测量与控制协会常务理事, 教育部学位与研究生教育发展研究中心通讯评议专家等近十余项学术兼职。

主要从事感知与测控技术方面的研究。近年来以项目负责人和分项目负责人身份完成了国家 863 计划、国家科技支撑、国家自然科学基金、国防基础科研、总装预研等项目二十余项。以第一作者或通讯作者发表 SCI/EI 收录论文 70 余篇, 以第一发明人申请国家/国防发明专利 50 余项, 获得发明专利授权 21 项; 主编/参编著作 4 部、译著 1 部。获省部级科技进步奖一等奖 1 项, 三等奖 2 项, 并多次在国内外学术会议上作特邀报告和分组报告。



郝创博(1989 -), 男, 博士, 主要研究方向为无线传感器网络技术。



《宇航计测技术》2019 年第 3 期目次

原子层沉积纳米钝化层薄膜厚度测量技术研究
 基于激光反射层析法的换能器声场测量技术研究
 取样示波器瞬态响应的光电校准技术研究
 微米尺度格栅标准样片的质量参数评价
 同心锥形 TEM 室结构设计及性能研究
 谐波阻抗调配器校准技术研究
 基于单直角电桥的交流电阻校准装置测量不确定度评定
 激光目标模拟器现场功率校准装置的设计
 70 ~ 100 MeV 准单能中子参考辐射场设计
 人工智能在矢量推力现场动态校准中的应用研究
 100 kHz ~ 1 MHz 频率范围激光法与互易法校准水听器

灵敏度比较
 基于室内无线模型的穿墙损耗校正
 某控制组合电磁发射超标分析及解决方案
 基于 CCD 自动巡线智能小车设计
 一种综合孔径微波辐射计成像处理算法
 基于 MATLAB 的间断照射雷达信号仿真技术研究
 一种基于系留无人机的图像遥测系统设计
 卫星导航终端设备在真实卫星信号下的动态定位精度测试方法研究
 LaBr3(Ce) 元件破损在线监测关键核素确定及效率校准
 信息安全绩效测量体系研究