

滑窗正态记分检验的传感网络数据融合方法*

朱铭雅^{1,2}, 舒坚^{1,3}, 刘琳岚^{1,2+}

¹(南昌航空大学 物联网技术研究所, 江西 南昌 330063)

²(南昌航空大学 信息工程学院, 江西 南昌 330063)

³(南昌航空大学 软件学院, 江西 南昌 330063)

Normal Score Test Data Fusion Method Based on Sliding Window for Wireless Sensor Networks

ZHU Ming-Ya^{1,2}, SHU Jian^{1,3}, LIU Lin-Lan^{1,2+}

¹(Internet of Things Technology Institute, Nanchang Hangkong University, Nanchang 330063, China)

²(School of Information Engineering, Nanchang Hangkong University, Nanchang 330063, China)

³(School of Software, Nanchang Hangkong University, Nanchang 330063, China)

+ Corresponding author: E-mail: linda_cn68@yahoo.com

Zhu MY, Shu J, Liu LL. Normal score test data fusion method based on sliding window for wireless sensor networks. Journal of Software, 2012, 23(Suppl. (1)): 9-16 (in Chinese). <http://www.jos.org.cn/1000-9825/12002.htm>

Abstract: Due to the declining accuracy caused by noise in wireless sensor networks, a data fusion method called normal score test, based on sliding window, is proposed. With this method, a measured data sample in a cycle is taken as a sliding window, and a double sample normal score test is employed into samples. Furthermore, a relationship matrix is created to maximize the adjacent subgraph. An estimated fusion value is achieved in terms of weighting and averaging its vertices. To some extent, this method makes good use of the character of a normal score test and restrains the impact noise on fusion. Simulation results demonstrate the effectiveness of proposed method.

Key words: wireless sensor network; sliding window; normal score test; data fusion

摘要: 针对无线传感器网络中, 节点受噪声影响导致测量精度下降的问题, 提出一种基于滑动窗口内正态记分检验的同质感知数据融合方法. 在节点内设置滑动窗口, 存储一个周期内的测量值样本数据, 并对样本数据进行双样本正态记分检验, 从而建立关系矩阵, 根据关系矩阵获得最大邻接子图, 加权平均子图顶点得到融合估计值. 该方法利用了正态记分检验良好的小样本检验特性, 在一定程度上抑制了噪声对融合结果的影响. 仿真结果表明了该方法的有效性.

关键词: 无线传感器网络; 滑动窗口; 正态记分检验; 数据融合

无线传感器网络(wireless sensor networks, WSNs)由大量传感器节点组成, 通过无线通信方式形成一个多跳的自组织网络, 可以协作地感知、采集和处理网络覆盖区域中感知对象的信息, 并发送给观测者^[1]. 传感器节点

* 基金项目: 国家自然科学基金(1262020); 江西省科技支撑重点项目(2009BGA01000)

收稿时间: 2012-05-05; 定稿时间: 2012-08-17

由电池供电,一般部署于环境恶劣、甚至人员无法到达的区域,使得更换电池为其补充能量变得不现实,因此需要对采样数据融合处理,降低冗余度,降低能耗,从而延长网络寿命^[2].WSNs 数据融合与传统数据融合不同,主要表现为节点芯片处理能力较弱,无线通信不稳定^[3].因此,计算复杂的融合方法不适合 WSNs.再者,因节点内部器件老化或受外界温度、压强、电磁辐射等突发性强干扰^[4],也会造成采集数据中含有噪声,一种良好的数据融合方法可以大大减小噪声对融合结果的影响.

目前,WSNs 数据融合方法主要有:算术平均法、加权平均法、Bays 估计法^[5]、分批估计法、卡尔曼滤波法^[6]等.WSNs 因其传感节点易受环境干扰而出现故障或失效的现象,直接运用简单的算术平均法或加权平均法^[7]常常会出现融合值与真实值偏离较大的现象.为了获得更准确的融合结果,文献[8]提出一种分批估计法,将各传感器均分成两批,分别计算每批的样本均值和样本方差后再采用方差加权方法进行融合处理,该方法在一定程度上节省了能量,但其实质仍属于加权平均法的范畴,未对采样数据噪声进行处理.为此需要探索具有去噪功能的数据融合方法.文献[9]提出卡尔曼波能大大减少计算量并获得良好的滤波效果,但当某个传感器受到外界强干扰或发生故障时,算法缺乏容错性.文献[10-12]中都提到窗口内的中值滤波能够有效地细化脉冲噪声,具有消除突发性强干扰,增强容错性的功能.文献[13]则针对不同噪声变化下窗长的取值给出了一定的阐述,但未给出具体的公式和证明.文献[14]则针对传感器噪声突变的特点,提出一种自适应窗长数据融合方法,该方法通过检测传感器数据中噪声的变化,将噪声的估计曲线分为平滑段和突变段,并对不同的曲线段运用不同长度的窗口对多传感器数据融合值进行修正,提高了融合精度.在文献[13,14]的基础上,本文提出一种滑动窗口内正态记分检验的同类传感器感知数据融合方法.

1 问题描述

如图 1 所示,WSNs 按地域划分成若干个独立的无线自组织网络监测区域,每个无线自组织网络监测区域以一个簇首节点为中心,散布在每个监测区域内的感知节点通过无线通信方式定期地将采样数据上传簇首节点,簇首节点再将收集到的信息定期上报给 sink 节点,由 sink 节点作数据级的同类传感器数据融合.在一个无线自组织网络监测区域同一类由 n 个传感器组成的多传感器系统,设定各传感器以相同的采样速率 $1/T$ 对同一环境特征(如环境湿度)进行采样,那么 T 就是系统的融合周期.

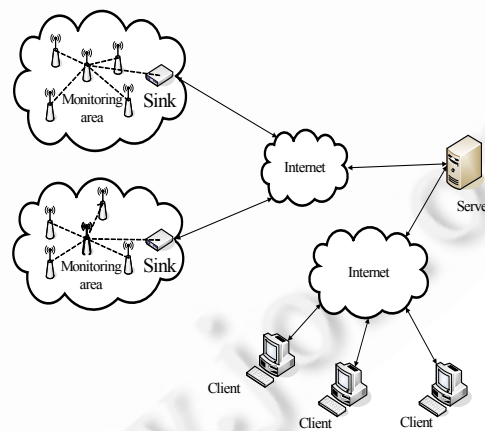


Fig.1 Environmental monitoring diagram

图 1 环境监测示意图

2 算法描述

在每个感知节点内设置滑动窗口,存放一个周期的测量值样本数据,采用双样本正态记分检验法对滑动窗口内样本数据进行检验并建立关系矩阵,根据关系矩阵获得传感器节点窗口内样本中位数相互支持的最大邻接子

图,加权平均子图顶点得到融合估计值.

2.1 滑动窗口内的正态记分检验

对两两传感器滑动样本数据进行正态记分检验.假定两样本(第 i 个传感器和第 j 个传感器滑动样本) X_1, \dots, X_n 和 Y_1, \dots, Y_n 分别来自样本中心为 M_X 和 M_Y 的总体.作如下假设:

$$\begin{aligned} H_0: M_X &= M_Y \\ H_1: M_X &\neq M_Y \end{aligned} \quad (1)$$

首先把两样本混合起来,并按升幂排列,把每一个观测值在混合样本中的秩 r 替换为 $r/(2n+1)$ 个标准正态分位点,计算某一个样本(哪一个都可以)的总正态记分 T ,然后根据统计量公式 $Z=T/S$ 得到两样本的统计量 Z .设第 i 个传感器所在样本的滑动窗口大小为 N ,其中 $k=1,2,\dots,N$,按照正态分布来定义记分函数,为了使 $a_N^+(k) \geq 0$,令记分函数为

$$a_N^+(k) = \Phi^{-1}\left(\frac{N+1+k}{2N+2}\right) = \Phi^{-1}\left[\frac{1}{2}\left(1 + \frac{k}{N+1}\right)\right] \quad (2)$$

相应的线性符号秩统计量的第 k 项为

$$s_k = \sum_{i=1}^n a_N^+(r_k) \text{sign}(X_k - M_X) = \Phi^{-1}\left[\frac{1}{2}\left(1 + \frac{r_k}{N+1}\right)\right] \text{sign}(X_k - M_X) \quad (3)$$

$$S_N = \sum_{k=1}^n s_k^2 \quad (4)$$

假设 $X_k - M_X$ 独立且对称分布, $E(S_N) = 0$, $\text{Var}(S_N) = \sum_{i=1}^n s_k^2$,由此,把 S_N 标准化,得到单样本(第 i 个传感器滑动窗口样本)位置的正态记分检验统计量:

$$T_i = \frac{S_N - E(S_N)}{\sqrt{\text{Var}(S_N)}} = \frac{S_N}{\sqrt{\sum_{i=1}^n s_k^2}} \quad (5)$$

根据统计量公式 $Z=T/S$ 得到两样本的统计量 Z :

$$Z = Z_{ij} = Z_{ji} = T_i / S \quad (6)$$

其中, $S^2 = \frac{N^2 \sum_k \omega_k^2}{2N(2N-1)}$; $\omega_k = \Phi^{-1}[k/N + N + 1]$; $k=1, \dots, 2N$,可计算出观测显著水平 p 值:

$$p = p_{ji} = p_{ij} = 1 - 2\Phi^{-1}(Z_{ij}) \quad (7)$$

2.2 关系矩阵的建立和异常传感器的剔除

对两两传感器样本数据进行双边检验,选取某一显著性水平阈值 α ,当传感器 i 与 j 所在样本构成的检验统计量 $p \leq \alpha$ 时,接受 H_0 ,在相应的关系矩阵中,元素 $r_{ij} = r_{ji} = 1$,反之 $p \geq \alpha$,拒绝 H_0 ,接受 H_1 ,在相应的关系矩阵中,元素 $r_{ij} = r_{ji} = 0$,根据下式:

$$r_{ij} = \begin{cases} 1, & p_{ij} < \alpha \\ 0, & p_{ij} \geq \alpha \end{cases} \quad (8)$$

得到关系矩阵:

$$R_{n \times n} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix} \quad (9)$$

设各传感器滑动窗口内样本的中位数为图 G 的顶点,关系矩阵 $R_{n \times n}$ 作为图 G 的邻接矩阵,可绘制出各传感器滑动窗口内样本的中位数的关系图.依照文献[15]求解图 G 的最大完全子图,即最大相互支持的传感器集合

A,其他传感器则组成异常传感器集合B,加权平均A中数据得到融合估计值.本文方法假设正常传感器占有率大于1/2,否则,可能错误地将异常传感器判断为正常传感器,而使得融合结果偏差较大.

2.3 权值的确定

根据测量模型 $z(k) = x(k) + \eta(k)$ ^[4].假设对 n 个正常传感器测量值进行加权估计,对应权值为 W_1, W_2, \dots, W_n , 则第 k 次测量的估计量为

$$\hat{z}_g(k) = \sum_{i=1}^n w_i z_i(k) \quad (10)$$

该估计量为无偏估计量,需满足以下条件:

$$E(\hat{z}_g(k)) = x(k) \quad (11)$$

由式(10)得到:

$$E\left[\sum_{i=1}^n w_i z_i(k)\right] = x(k) \quad (12)$$

由测量模型得到:

$$E\left\{\sum_{i=1}^n w_i [x(k) + \eta_i(k)]\right\} = x(k) \quad (13)$$

化简得到:

$$\sum_{i=1}^n w_i E[x(k)] + \sum_{i=1}^n w_i E[\eta_i(k)] = x(k) \quad (14)$$

则 $E[\eta_i(k)] = 0$ 且 $E[x(k)] = x(k)$, 因此可以得到 $\sum_{i=1}^n w_i = 1$. 该估计量为均方差最小估计量,需满足 $D\{[\hat{z}_g(k) - x(k)]^2\}$ 取最小值.

由于 $z_1(k), z_2(k), \dots, z_n(k)$ 相互独立且 $\sum_{i=1}^n w_i = 1$, 可得:

$$D\{[\hat{z}_g(k) - x(k)]^2\} = \sum_{i=1}^n w_i^2 [\sigma^2(k)] \quad (15)$$

采用拉格朗日条件求极值方法可得:

$$W_i = \frac{1/[\sigma^2(k)]}{\sum_{j=1}^n 1/[\sigma^2(k)]} \quad (16)$$

其中, $\sigma^2(k)$ 为传感器 i 的综合噪声方差.

各传感器第 k 次测量的噪声方差 $\sigma^2(k)$ 未知,采用滑动窗口的样本方差代替 $\sigma^2(k)$, 得到:

$$W_i = \frac{1/S_i^{*2}(k)}{\sum_{j=1}^n 1/S_j^{*2}(k)} \quad (17)$$

3 仿真实验

3.1 仿真说明

实验在 MATLAB 平台下进行,假设只对单个无线自组织监测区域进行融合,区域内部署 200 个节点,在 sink 节点进行融合,节点内滑动窗口大小为 N ,最优显著性水平阈值为 α_{best} ,测量次数为 n_{total} ,经验误判率为 $P=0.1$. 在测量模型 $z_i(k) = x_i(k) + \eta_i(k)$ 中, $z_i(k)$ 为传感器 i 的第 k 次测量值, $x_i(k)$ 为传感器 i 的第 k 次实际值, $\eta_i(k)$ 为传感器 i 的第 k 次综合噪声,各个传感器之间的综合噪声相互独立.内部噪声采用均值非零且方差随机的高斯白噪

模拟,外界噪声采用均值为0且方差随机的高斯白噪模拟,考虑采样周期为1s,为体现外界噪声的时变特性,设外界噪声每隔 R_y 秒变化1次,目标对象实际值 $x_i(k)$ 随机产生且每隔 R_x 秒变化1次。

3.2 阈值和滑动窗口值的确定

设置不同显著性水平阈值 α 和不同窗口大小 N ,根据实验结果确定最优显著性水平阈值 α_{best} 和最优滑动窗口值 N_{best} 。图2为不同显著性水平阈值 α 和不同窗口长度 N 下的实验结果:

1) 当 $N \in [1,5]$ 时,均方误差都呈急剧下降趋势,原因在于滑动窗口值接近真实值变化率 R_x 和综合噪声变化率 R_y ,此外,窗口内数据元素较少且随着窗口的增大,信息量也增大。

2) 当 $N \in [5,11]$ 时,均方误差趋于平稳且可取最小值,原因在于窗口值在 R_x 和 R_y 附近且此时信息量较充裕,因此 α 在0.1附近时,可取得最佳融合效果。

3) 当 $N \in [11,30]$ 时,均方误差趋于平稳并缓慢增加,原因在于窗口值均大于 R_x 和 R_y ,信息量已饱和,随着信息量的增加,均方误差趋于平稳,但由于水平阈值 α 增大,误判率 P 上升,融合效果下降。综上所述,最优显著性水平阈值 α_{best} 取值范围为 $[0.05,0.15]$,且对应的最优滑动窗口长度 $N_{best}=6$ 。

3.3 实验结果及分析

实验以总均方误差 $\delta_{total} = \sum_{k=1}^n |x(k) - \hat{z}(k)|$ 作为衡量标准。设置感知节点数 $m=200$,测量次数 $n_{total}=100$,在最优滑动窗口值 $N_{best}=6$ 和 $\alpha_{best}=0.1$ 条件下的实验结果与算术平均法、分批估计法、自适应加权法这3种数据融合方法进行比较。总均方差如图3所示,均方误差如图4(a)和图4(b)所示。

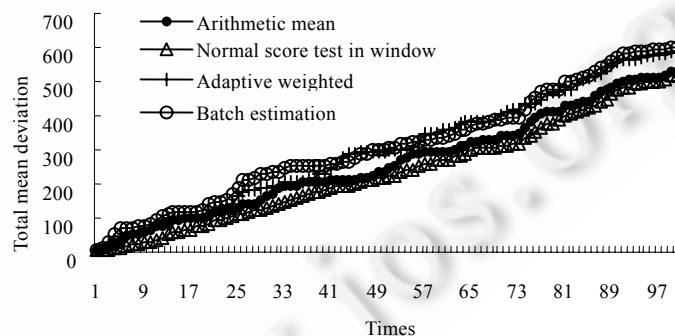


Fig.3 Total mean deviation comparison in different methods

图3 各方法总均方差比较

(1) 算术平均法是最简单的方法,算法复杂度最低,对所有传感器节点的数据进行平均运算,得到融合估计值。没有对采样数据进行检验分析,也没有计算每个传感器的权值,只是简单的平均运算,通过平均运算可能会掩盖某些传感器稳定性差给融合带来的影响,而使得该方法在精度方面存在不稳定性。

(2) 分批估计法将所有传感器节点的测量数据均分成样本集合 A 和 B ,并分别计算集合 A 和 B 的样本均值和样本方差,得到融合估计值。该方法的算法复杂度也不高,也只是简单地对传感器进行分批处理,利用各批次

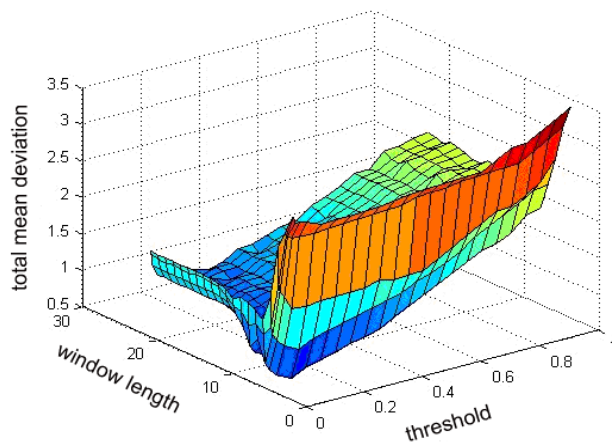


Fig.2 Determination of the window size and threshold

图2 窗口值和阈值的确定

样本方差的倒数作为权值及样本均值进行简单加权运算,忽略了各传感器噪声不一的特性,实验结果表现出其误差也较大.

(3) 传统自适应滑窗法计算各传感器节点的滑动窗口中测量值的样本方差,然后根据假设检验建立关系矩阵,最后采用方差加权方法进行融合处理,得到融合估计值.传统自适应滑窗法一般采用基于正态分布的假设检验方法,该检验方法理论上要求数据量越大越好,对于传感器网络用于环境监测,由于上传数据的周期为 1s,不可能采集到大量数据.本文做了这两种方法不同样本数量与均方差比较实验,结果如图 5 所示,证实了正态记分检验方法具有良好的小样本特性.

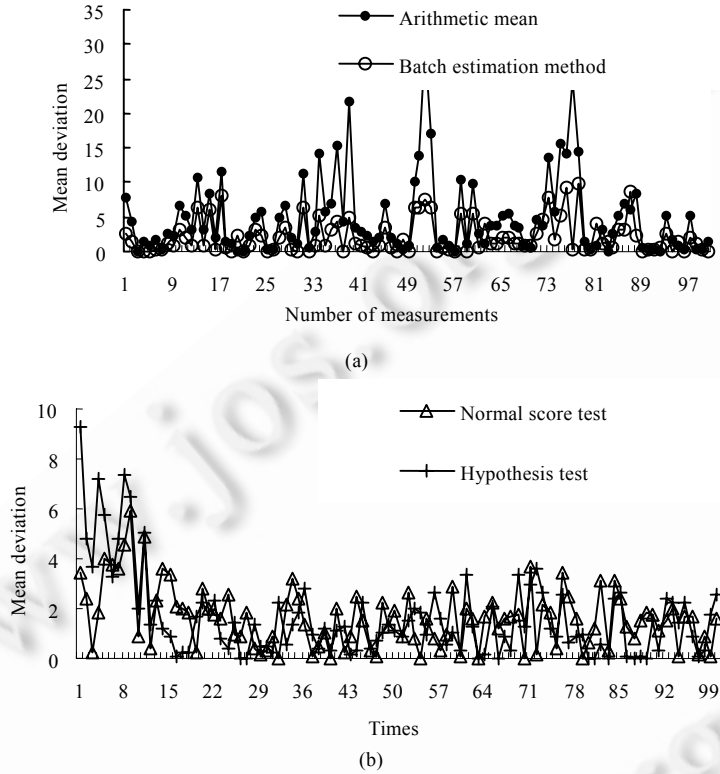


Fig.4 Mean deviation comparison in different methods
图 4 各方法均方差比较

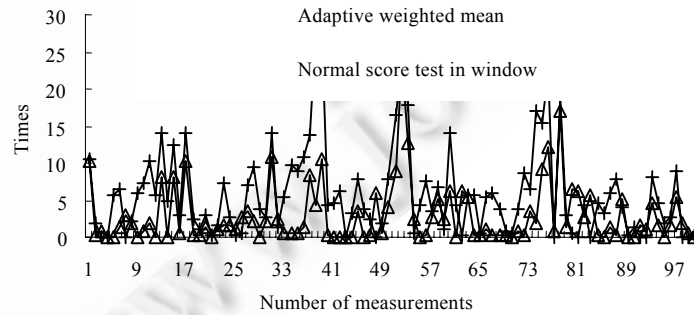


Fig.5 Comparison of test methods
图 5 两种检验方法的比较

综上所述可得,本方法融合效果最佳,原因在于:① 滑动窗口最优值为小样本数据,正态记分检验法具有较

好的小样本检验特性,使融合信息定位更准确;② 取各传感器滑动窗口内的样本中位数作为邻接图的顶点,借鉴中值滤波的去噪思想,抑制部分脉冲噪声对融合结果的影响;③ 根据关系矩阵求得滑动窗口内样本相互支持的最大子图,有效地剔除失效或故障的传感器测量数据。②和③可从图 4(a)和图 4(b)中看出,当传感器受到脉冲噪声或者出现故障和失效的传感器时,其他 3 种方法没有很好的容错能力,出现了均方误差的较大波动,本文方法采用了中值平滑原理并剔除异常传感器,均方误差没有出现较大波动。综合上述原因,本文方法提高了融合数据的精度,但在算法复杂度上,较算术平均法和分批估计法大不少,与传统自适应滑动窗方差加权法差不多。本文方法的融合估计值更接近实际值,在上述方法中融合效果最佳。

4 结 论

在节点内设置滑动窗口,窗口内进行正态记分检验的数据融合方法,利用了正态记分检验良好的小样本特性,结合关系矩阵,以各传感器滑动窗口内样本的中位数建立邻接图,从而获得最大相互支持的传感器样本中位数集合,可滤除测量数据上的部分脉冲噪声,剔除故障或失效传感器的测量数据,提高参与加权平均的样本数据的可靠性,仿真结果表明,该方法是有效的。

References:

- [1] Rezaei Z, Mobinnejad S. Energy saving in wireless sensor networks. *Int'l Journal of Computer Science & Engineering Survey*, 2012,3(1):23-37. [doi:10.5121/ijcses.2012.3103]
- [2] Peng AP, Guo XS, Cai W, Xu XM. A data aggregation algorithm for clustering wireless sensor networks based on estimative scheme. *Chinese Journal of Sensors and Actuators*, 2011,24(1):128-133 (in Chinese with English abstract). [doi:10.3969/j.issn.1004-1699.2011.01]
- [3] Kong XS, Zhao DG, Wang DH, Zhang ZJ. The impact analysis of wireless sensor network in low-level channel. *Chinese Journal of Sensors and Actuators*, 2011,24(1):106-110 (in Chinese with English abstract). [doi:10.3969/j.issn.1004-1699.2011.01.022]
- [4] Duan ZS, Han CZ, Tao TF. Consistent multi-sensor data fusion based on nearest statistical distance. *Chinese Journal of Scientific Instrument*, 2005,26(5):478-481 (in Chinese with English abstract).
- [5] Zhang SK, Cui ZM, Gong SR, Sun Y, Fang W. A data fusion algorithm based on bayes sequential estimation for wireless sensor network. *Journal of Electronics & Information Technology*, 2009,31(3):716-721 (in Chinese with English abstract)
- [6] Olfati-Saber R. Distributed Kalman filter with embedded consensus filters. In: *Proc. of the 44th IEEE Conf. on Decision and Control, and the European Control Conf.* 2005. 8179-8184.
- [7] Miu XC, Qiu M, Mai HR. Study on data fusion algorithms based on parameter-estimation. *Transducer and Microsystem Technologies*, 2006,25(10):70-73 (in Chinese with English abstract).
- [8] Zhang LY, Zhang L, Li D. Grouping weighted fusion algorithm for multi-sensor and its optimal grouping principle. *Chinese Journal of Scientific Instrument*, 2008,29(1):200-205 (in Chinese with English abstract).
- [9] Wei XY, Miu XC. Research on spatial-temporal fusion for WSNs based on Kalman filter. *Transducer and Microsystem Technologies*, 2007,26(9):72-75 (in Chinese with English abstract).
- [10] Lin T-C. A new adaptive center weighted median filter for suppressing impulsive noise in images. *Information Sciences*, 2007, 177(4):1073-1087.
- [11] Wang X. Multiscale median filter for image denoising. In: *Proc. of the ICSP*. 2010. 2617-2620.
- [12] Zhang L, Chen ZQ, Gao WH, Kang KJ. Mean-Based fast median filter. *Journal of Tsinghua Univ. (Sci & Tech)*, 2004,44(9):1157-1159 (in Chinese with English abstract).
- [13] Xiao LY, Zeng C. A daptive second data fusion algorithm. *Transducer and Microsystem Technologies*, 2007,26(4):81-83 (in Chinese with English abstract).
- [14] Zhang Y, Jia M. The application of variance-estimation based on the adaptive window length in multi-sensor data fusion. *Chinese Journal of Sensors and Actuators*, 2008,21(8):1398-1401 (in Chinese with English abstract).
- [15] Konc J, Janžić D. An improved branch and bound algorithm for the maximum clique problem. *MATCH Commun. Math. Comput. Chem.*, 2007,58:569-590.

附中文参考文献:

- [2] 彭爱平,郭晓松,蔡伟,徐晓森.基于估计机制的分簇传感器网络数据融合算法.传感技术学报,2011,24(1):128-133. [doi: 10.3969/j.issn.1004-1699.2011.01]
- [3] 孔祥善,赵德光,王代华,张志杰.低空信道对无线传感器网络的影响分析.传感技术学报,2011,24(1):106-110. [doi: 10.3969/j.issn.1004-1699.2011.01]
- [4] 段战胜,韩崇昭,陶唐飞.基于最近统计距离的多传感器一致性数据融合.仪器仪表学报,2005,26(5):478-481.
- [5] 张书奎,崔志明,龚声蓉,孙涌,方巍.基于 Bayes 序贯估计的无线传感器网络数据融合算法.电子与信息学报,2009,31(3):716-721.
- [7] 缪惜春,丘敏,麦汉荣.基于参数估计的数据融合算法研究.传感器与微系统,2006,25(10):70-73.
- [8] 张立勇,张利,李丹.基于最优分组原则的多传感器分组加权融合算法.仪器仪表学报,2008,29(1):200-205.
- [9] 魏雪云,廖惜春.基于卡尔曼的无线传感器网络时空融合研究.传感器与微系统,2007,26(9):72-75.
- [12] 张丽,陈志强,高文焕,康克军.均值加速的快速中值滤波算法.清华大学学报:自然科学版,2004,44(9):1157-1159.
- [13] 肖龙远,曾超.一种自适应二次数据融合方法.传感器与微系统,2007,26(4):81-83.
- [14] 张怡,贾民平.自适应窗方差估计在多传感器数据融合中的应用.传感技术学报,2008,21(8):1398-1401.



朱铭雅(1986—),女,硕士,主要研究领域为无线传感器网络.



刘琳岚(1968—),女,教授,CCF 会员,主要研究领域为无线传感器网络.



舒坚(1964—),男,教授,CCF 高级会员,主要研究领域为无线传感器网络,软件工程.