一种基于滑动聚类的多传感器异步信息融合方法

姓名：岳高峰，学号：4121357003，班级：B2174

（西安交通大学电信学部网络空间安全学院 西安 710049 643375731@qq.com）

**摘 要：**传感器是智能检测和自动化装置中重要的部件组成，为了解决多传感器异步数据下的融合难题，我们提出了多源异步下一种创新的基于滑动聚类的多传感器信息融合方法，核心点是我们引入了K-Means聚类方法，主要通过对k取值进行自适应设计替代人为指定，创新地设计了自适应参数的聚类滤波核，形成在时空域上滑动聚合的多传感器信息融合方法。仿真实验验证了本文方法的正确性以及合理性，可以应对时间域上数据源异步的问题。

**关键词：**多传感器融合；多源异步；K-Means聚类；滑动聚合；自适应滤波核

# 0 引言

随着自动化程度的提高，传感器稳定性需求急剧增加，传统单一的传感器无法支持系统需求，尤其在自动驾驶领域中更为显著[1]。随着人工智能近年来的快速发展，多传感器信息融合（Multi-Sensor Information Fusion，MSIF）得到快速的发展，例如高级驾驶辅助系统[2]，这些系统逐渐发展变得更加强大并在越来越多的情况下控制着智能装置。

MSIF又可以叫做多源信息融合，起源于20世纪70年代，主要应用于军事领域。近年来对于MSDF的研究取得了一些突破性发展，并应用在民用行业，例如同步定位与导航[3]、无人机编队飞行[4]、雷达成像等[5]。现有的研究方法主要有基于神经网络、统计推理、估计融合等三类。黄衍标等人[6]于2016年提出了基于BP神经网络的多传感器融合方法，由于BP网络自身的一些缺陷，通常还需要辅助一些优化算法，例如蚁群算法等；Meng等人[7]提出了基于深度学习的数据融合方法研究，深度学习所具有的强表示能力是BP网络所不及的，这些都归属于基于神经网络的研究方法。此外还有基于统计推理的一类方法，林水生等人[8]设计的D-S融合检测方法，该方法对于解决多源数据融合具有明显效果；Pan等人[9]以支持向量机方法为基础，设计了一个基于多传感器数据融合的跌倒检测系统，验证了多传感器数据融合算法的有效性；Gong等人[10]提出了一种新方法，用改进的卷积神经网络加支持向量机（Convolutional Neural Network-Support Vector Machine, CNN-SVM）的方法，该方法通过引入SVM，改进了传统的卷积神经网络模型结构，不过这种方法计算量急剧增加，对普通嵌入式设备无法使用。最后一类是基于估计理论的数据融合，例如常用的状态估计方法是卡尔曼滤波器（Kalman Filter，KF）[11]，这种方法适用于线性系统，但现实世界中的大多数系统通常是非线性的，为了克服这个问题，大量学者研究了基于KF的改进型滤波器，如扩展卡尔曼滤波器（Extended Kalman Filter，EKF）[12]，无痕卡尔曼滤波器（Unscented Kalman Filter，UKF）[13]。

总体而言，在数据融合理论的研究型论文数量较少，上述方法都有独有的使用场景，共同特点是基于同步假设下，主要面临的问题是无法解决实际情况下的挑战，例如传感器信息的异步性或者丢包等情况，这些都严重影响数据融合结果，因此，需要设计一种高性能和高容错性的多传感器信息融合算法。本文基于分布式多传感器融合框架（如图1所示），提出了异步下基于滑动聚类的多传感器信息融合新方法，针对时间域上数据非同步的问题，我们引入了DBSCAN聚类方法，通过定义自适应滤波核形成在时空域上滑动聚合的信息融合方法。

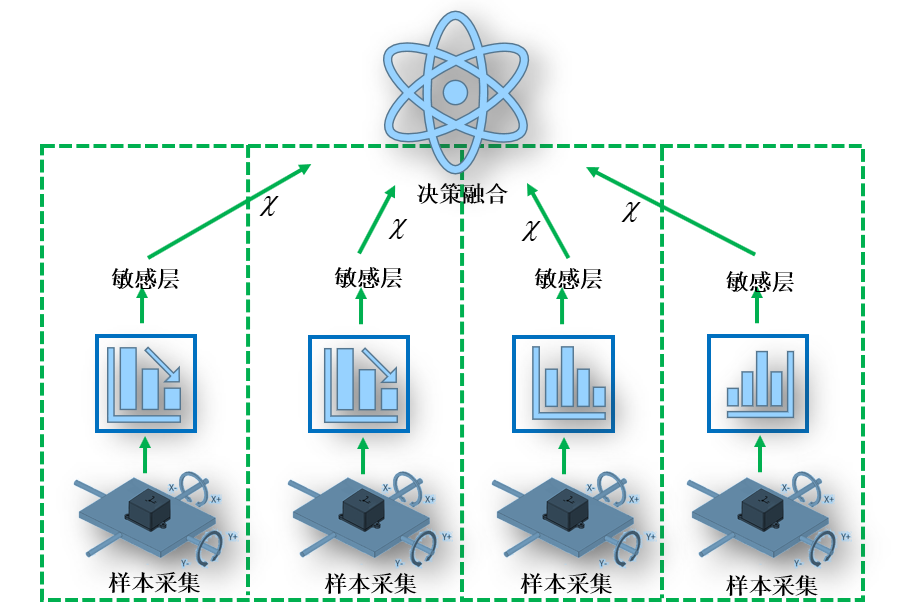


图1 分布式传感器融合架构

Fig.1 Distributed sensor fusion architecture

# 1 研究动机

与传统的这些方法不同，我们将使用聚类的方法进行异步补偿设置，在此之前我们先观察下面两组小实验，用以验证我们所提方法的必要性。

观察1：异步检测。我们设计一组数据收集实验，读取一个机器人启动时的四个传感器数据，并进行数据分析，对于多个传感器数据我们绘制图1，图中S1、S2、S3、S4代表四个传感器的启动时刻，如果是同步状态，则其应该是在一条竖直线上，而图中则不是，S2和S3距离虽然很近但也却不同步，可不可以忽略这个时间差，当然是可以的，不过这对于需求高精度的自动化装置中是不能接受的，因此需要一种容忍这个现象的新方法。

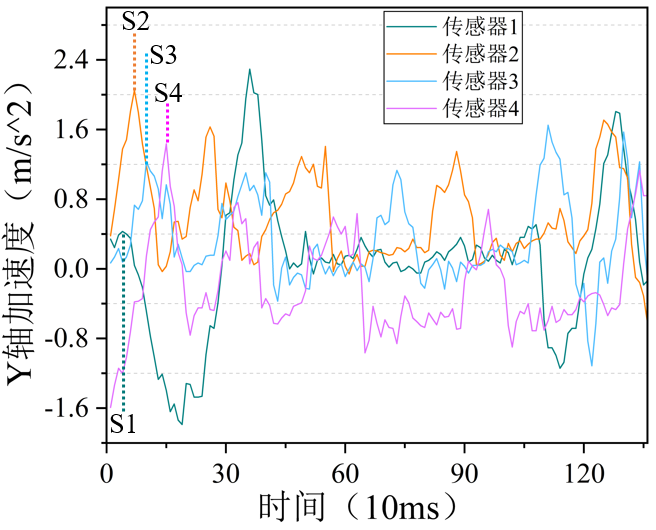


图2 多传感器异步采样

Fig.2 Multi-sensor asynchronous sampling

观察2：聚类分析。对于异步的解决，聚类是一种有效的方式，常见的核心聚类方法有基于划分方法的K-均值聚类（K-Means）[14]、基于密度的DBSCAN聚类[15]、基于层次的Agglomerative聚类[16]。为了验证这些方法对本文研究对象即传感器数据的效果，我们观察图3所示的结果。三种算法的聚类结果基本能够服从原始数据的分布，K均值聚类结果的每一类的时间长度基本相同，效果稳定，比较符合我们想要的自适应滤波核形状，方便处理数据中的噪声[17]，所以我们选用K-Means算法作为本文核心聚类方法。

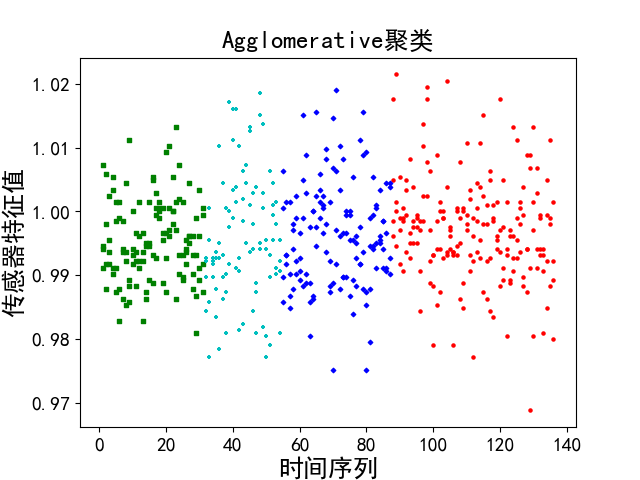
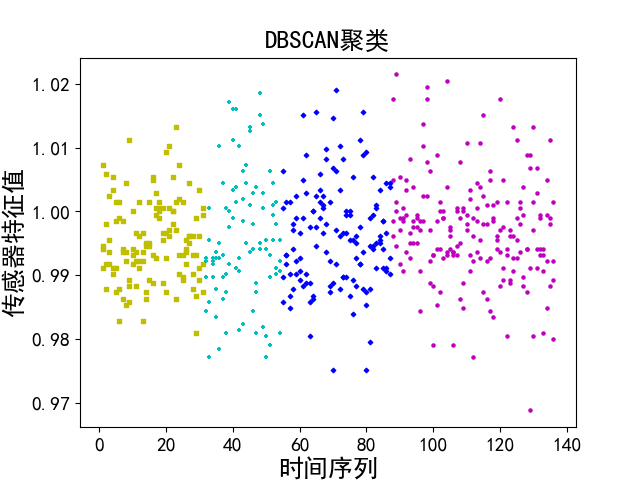
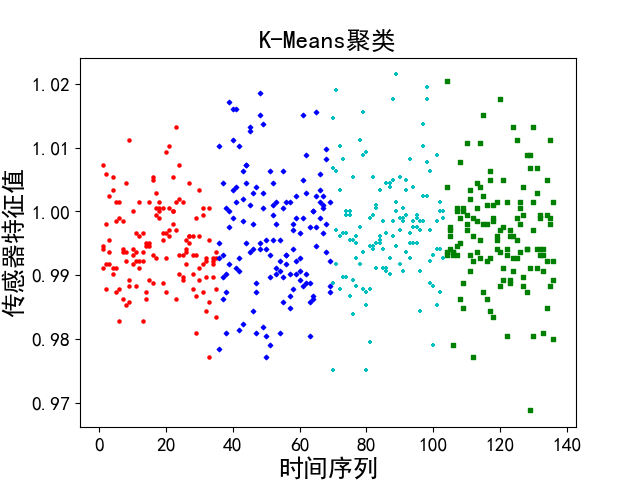
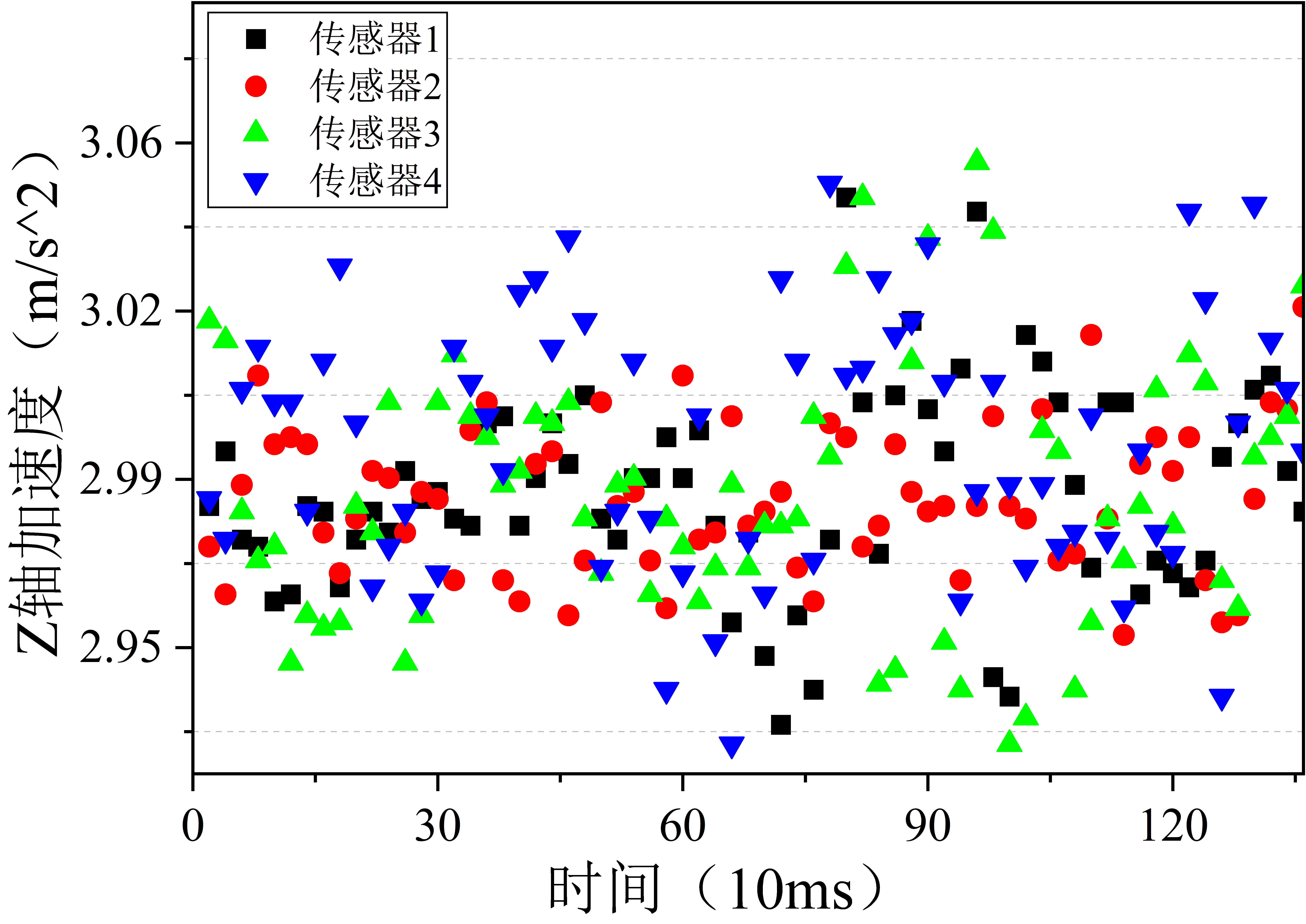


图3 聚类算法性能比较

Fig.3 Performance comparison of clustering algorithms

# 2 K-Means聚类原理

聚类是一种重要的无监督方法，聚类的目的是将一个数据集分成多个簇（或类），簇内相似度高，簇间相似度低。B.MacQueen在1967年提出的K-Means算法，是到目前为止用于科学和工业应用的诸多聚类算法中一种极有影响的技术[18]，一般来说，K-Means最小化每个样本点与其最近的聚类中心之间的欧几里得距离平方之和[19]，K-Means首先随机或手动选择初始聚类中心，然后将一个数据集划分为若干个聚类（一个数据点属于聚类中心离该数据点最近的聚类），并计算一个聚类的平均值作为聚类中心。K-Means重复地更新聚类中心和聚类，直到收敛。

让代表一个等待聚类的数据集，假设K-Means将其分为个聚类，表示第个簇的元素，代表类中的元素个数，我们有，聚类后的每个簇的中心为，k-means可以被描述为最小化下面的目标函数。

 （1）

 （2）

重复上述过程即可进行求解，而停止条件是迭代收敛或者达到稳定不变的误差。K-Means是聚类方法中一个基本的方法，常常采用误差平方和代价函数为聚类准则函数，K-Means的主要优点在于其原理比较简单，收敛速度快，聚类效果较优。

# 3 模型建立与算法设计

3.1 全局模型

我们分析数据的基本结构，从宏观来看，我们的多传感器序列数据是三维的，即使传感器数据的采集时不同步的，我们可以选择相邻一些时刻的值进行聚类分析再进行融合，如图4所示。

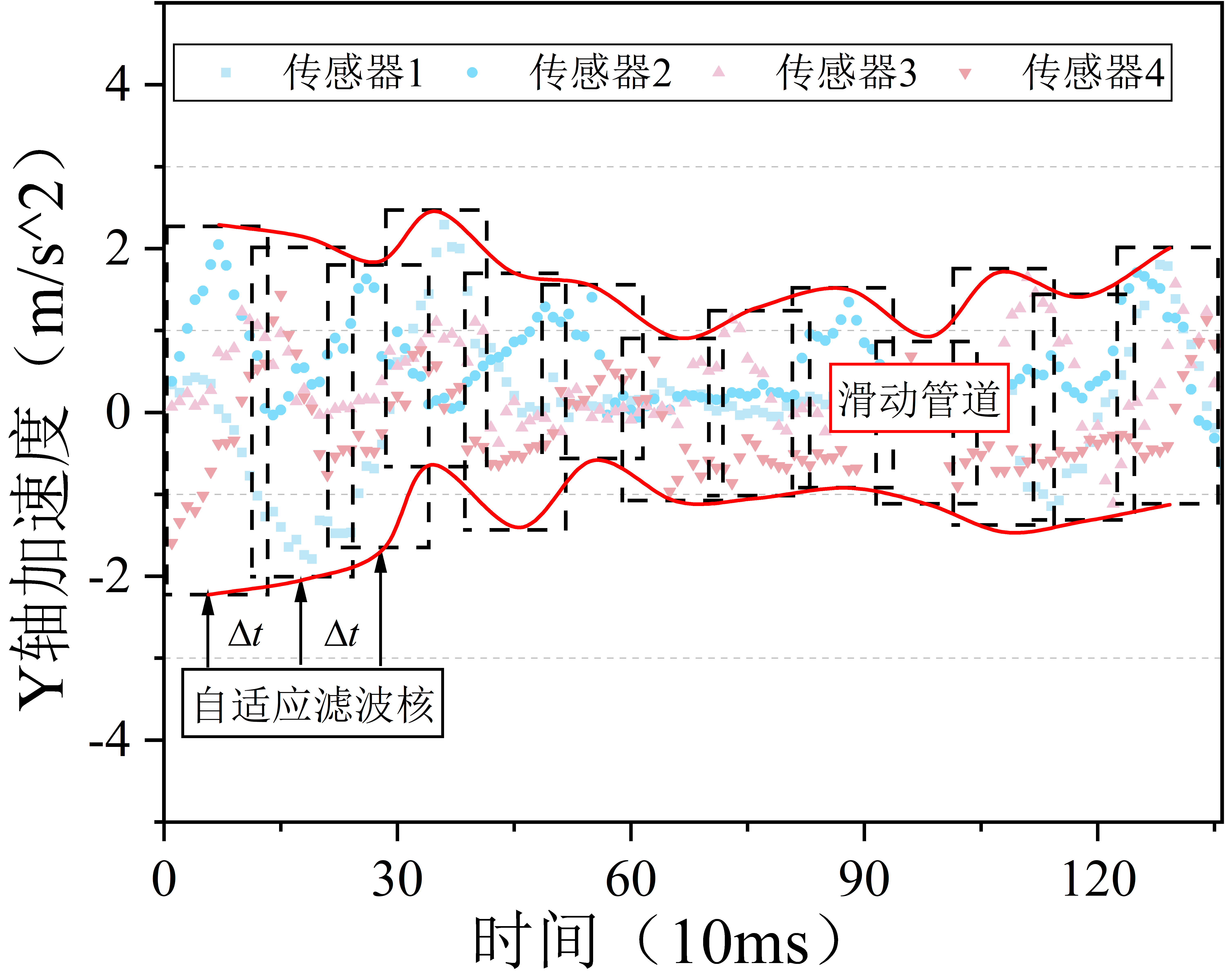


图4 算法示意图

Fig.1 Algorithm schematic

假设我们只考虑单个自适应滤波核中的时刻集合，定义代表该核中的传感器数据集合，代表该滤波核中第个传感器的序列数据，。假设聚类结果为类，每一类的中心为，则滤波核中聚类中心集合的输出结果为:

 (3)

式中为滤波核所占用的时间差，通过加权均值的方法可计算得到最终结果。

 (4)

式中为权值向量，可以为传感器出厂精确度，为中心位置向量，不过此处已经转变为，因此我们通过先进性聚类分析得到多传感器数据的基本分布特征，也得到了单个时刻最终的融合结果，通过滑动自适应聚类滤波核可实现实时的多传感器信息聚合并且形成滑动管道。

3.2 自适应聚类滤波核设计

在上述模型中我们采用了聚类滤波核，但是这里存在几个参数需要确定，如滤波核高度、中心位置、聚类个数等，如图5中所示。

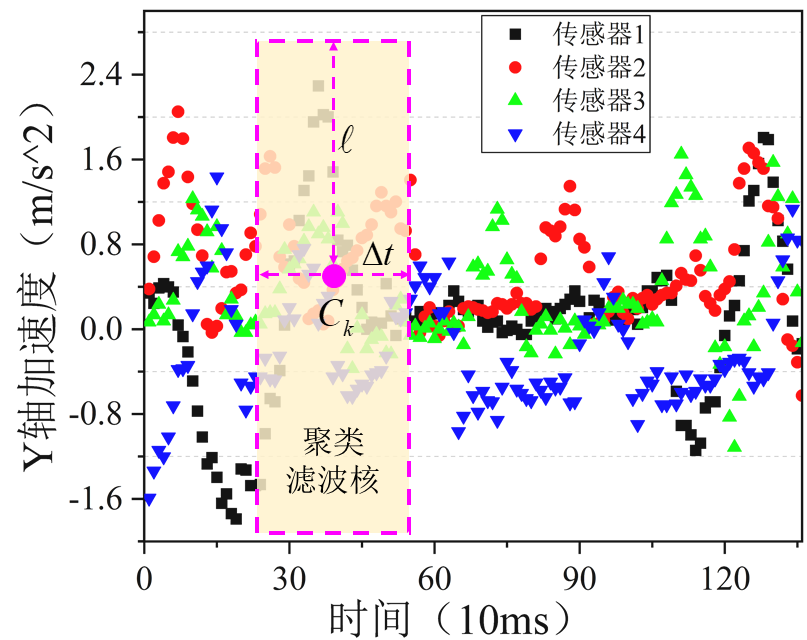


图5 滤波核示意图

Fig.5 Filtering kernel schematic

为了提高实用性，可以采用自适应设计，减少初始化参数赋值。对于可以根据实际情况进行变化，当然太大或者太小都会影响实际效果。和可以由下式计算。

 （5）

的计算我们可以引入皮尔逊相关系数[20],这样总是可以找到每个聚类簇的合理较差位置，任意两个簇的相关系数可以由下式计算：

（6）

其中，代表簇向量的数学期望。每个簇都会有对应的三个相关系数，我们挑选聚类结果最为相关的一组得，将进行归一化可得到最终的的解析式，如下式：

 （7）

式中为单个聚类簇的元素的均值。

实现实时聚类方法中的值是一个比较复杂的问题，我们给出一种简易的快速的判定法则，主要利用曲线拟合完成。首先我们将单个时刻的滤波核数据投影到单个特征轴（Y轴）形成集合，并划分为个组并进行组间元素数量求和，其数学模型如下。

 （8）

 （9）

式中，因此我们得到了个标量组成的集合。

（10）

利用最大可能准则（Maximal Feasible Criterion，MFC）对采用进行选择，即。假设传感器其噪声点满足泊松分布[21-22]，通过求解概率密度值的逆函数值，而这个值正好对应了K-Means所需的k值。

表1 多传感器融合伪代码

Table.1 Pseudo-code of the multi-sensor fusion

|  |
| --- |
| 伪代码: 多传感器融合算法 |
| **Initialization**: ，，， |
| **WHILE**  计算滤波核高度  计算中心位置  计算和k值  **FOR** |
| 使用K-Means计算  通过滤波核计算  **WHILE**  **END** |
| **OUTPUT** |
|  |
| **END** |

# 4 实验结果与分析

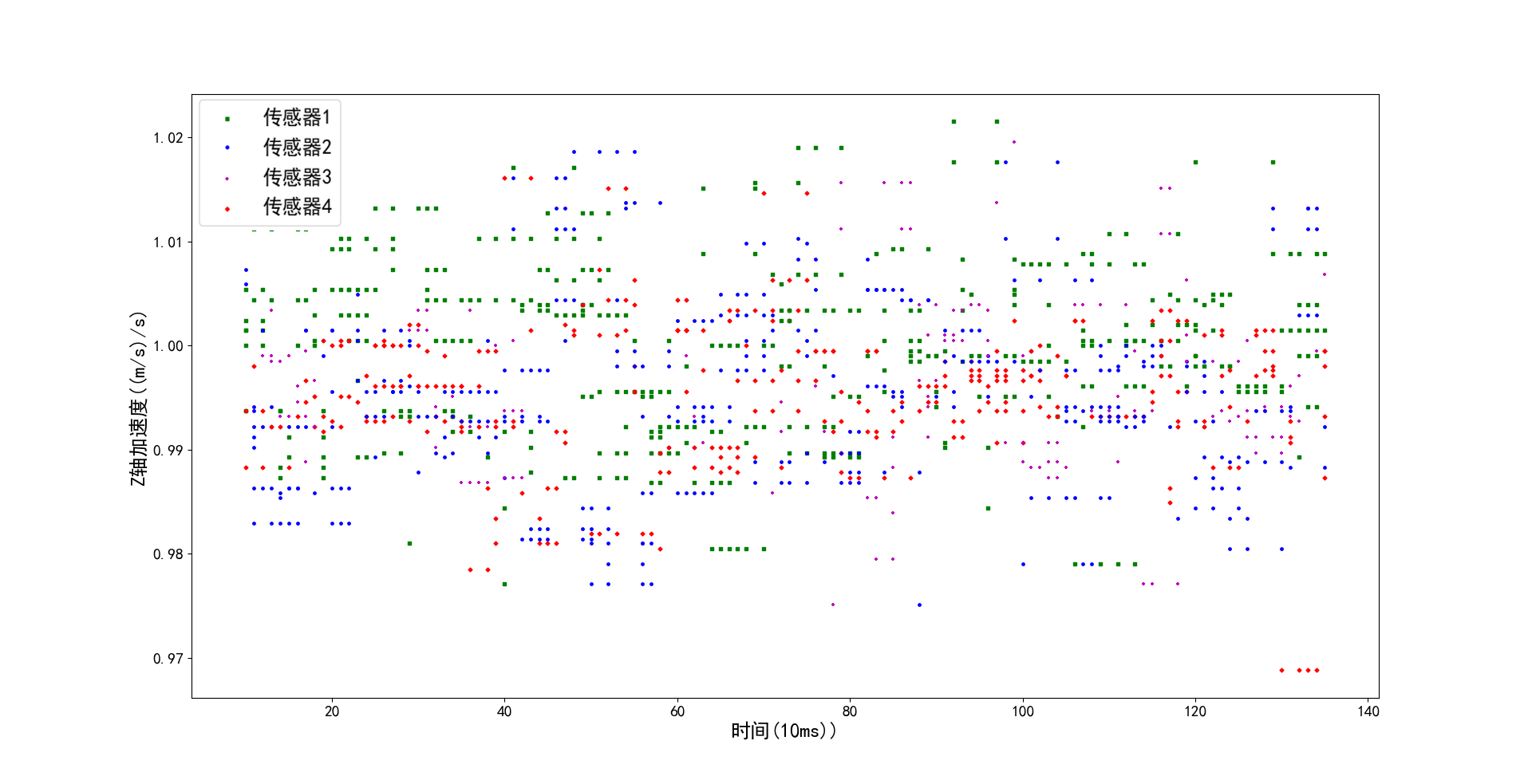
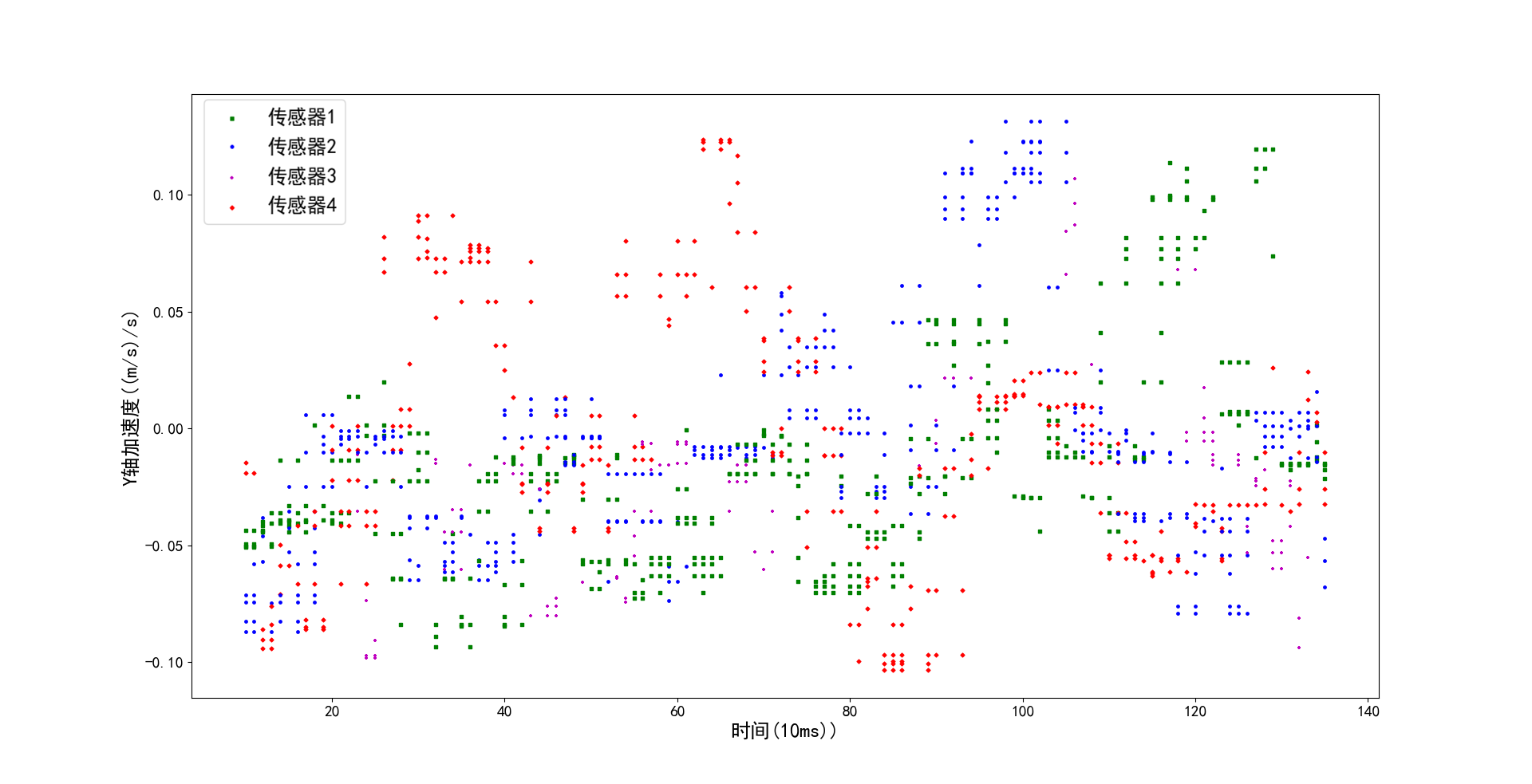
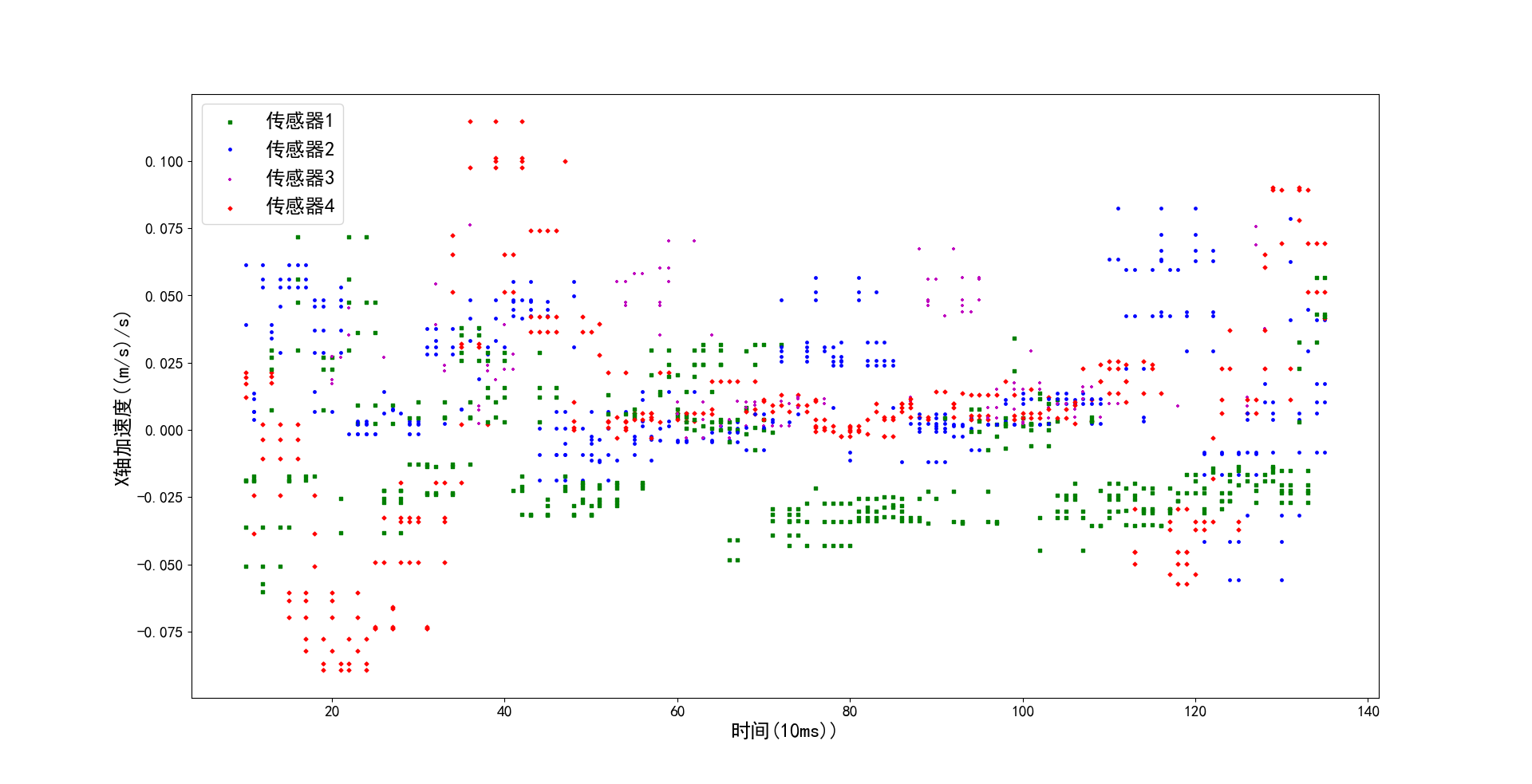
4.1实验平台

本文设计了计算机仿真实验，主要验证算法的合理性。仿真实验所用计算机基本配置如下，CPU为Intel(R) Xeon CPU E5-2680 v2 @ 2.80GHz；机带RAM为128G，操作系统为Windows 10专业版；编程语言为Python（版本3.9.7），开发环境为VS Code。

我们采用了四个MPU6500型号的陀螺仪传感器，采集了三个轴的加速度值。

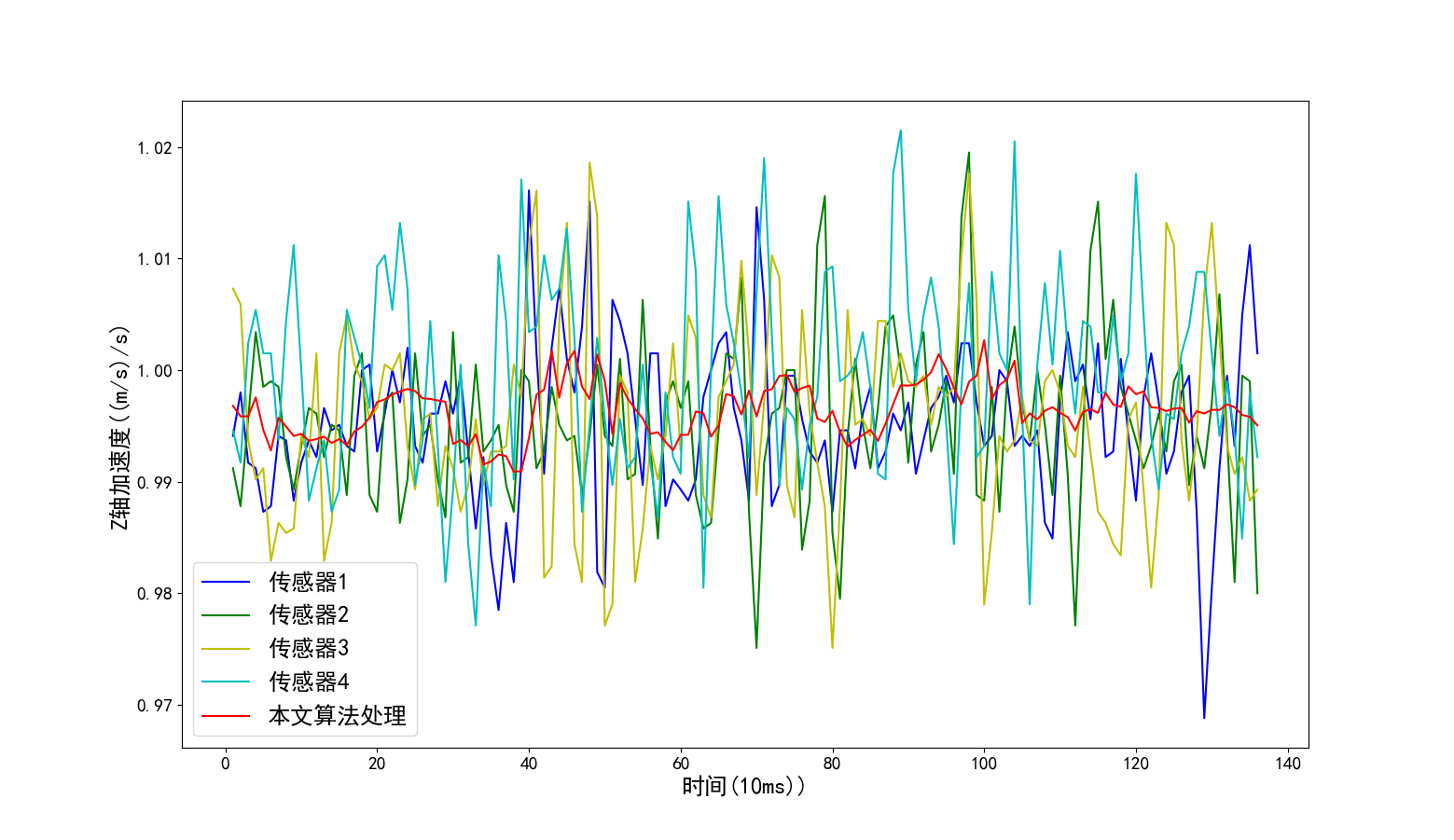
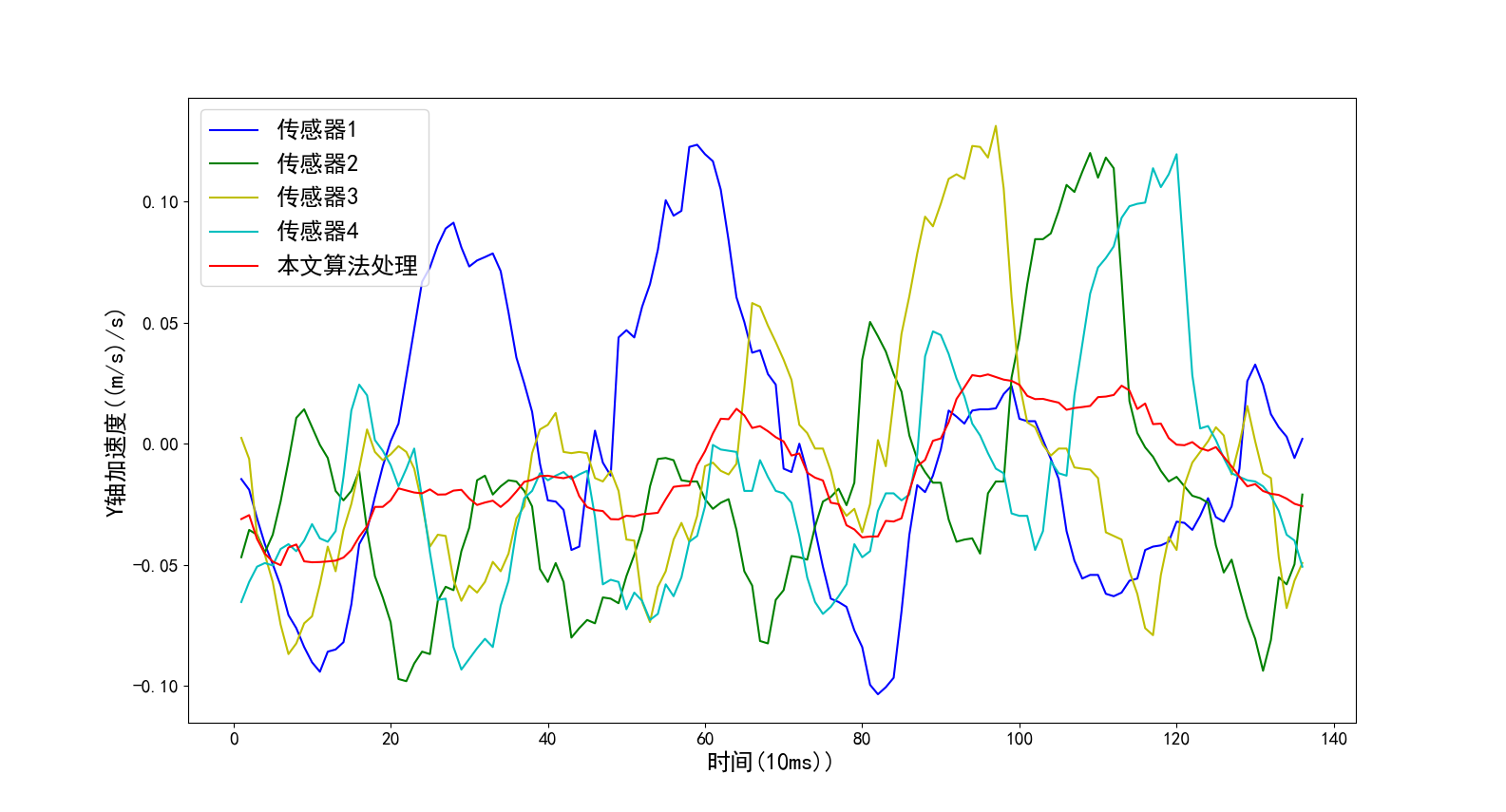
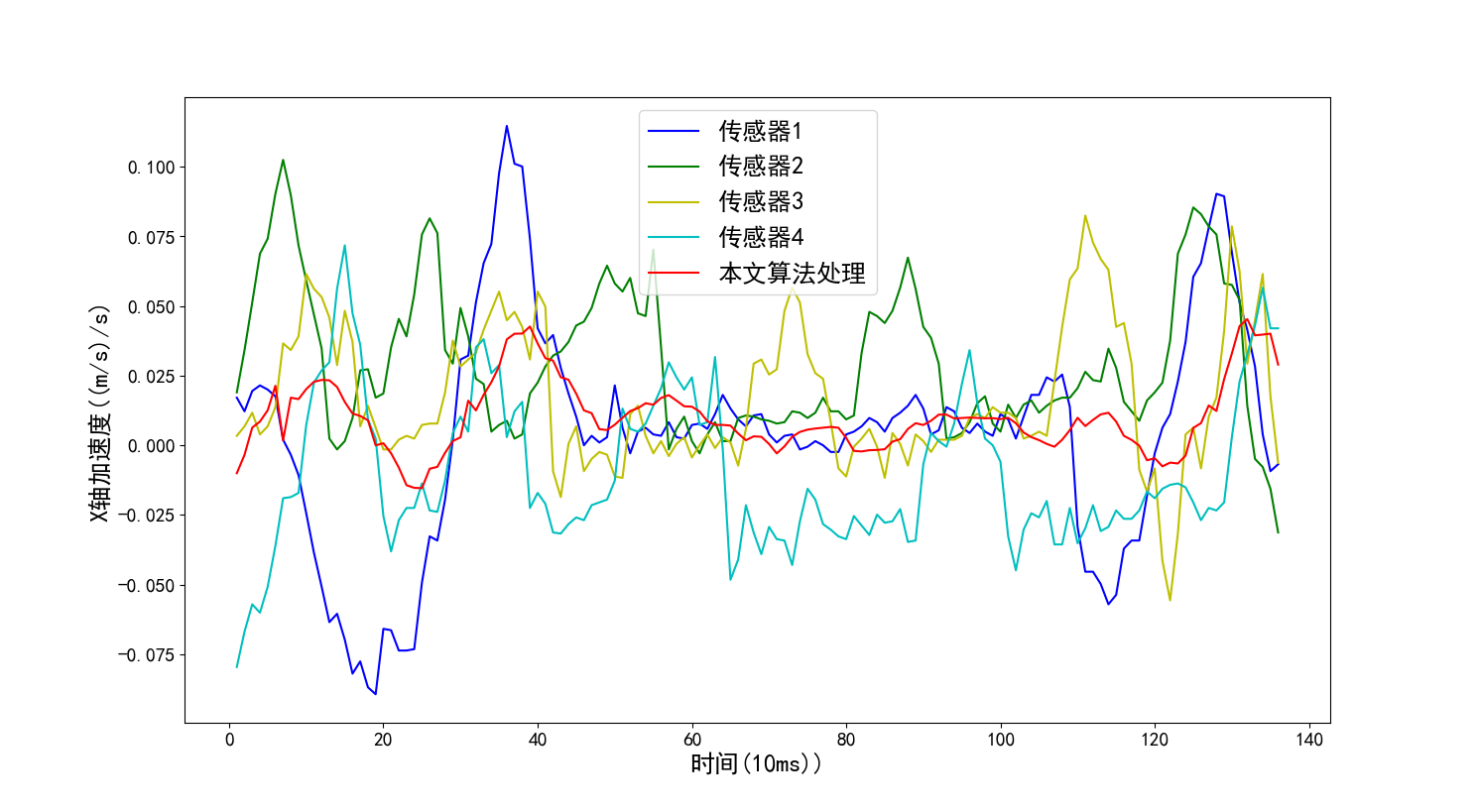
4.2实验结果

本文设计了计算机仿真实验，主要验证算法的合理性以及正确性。本文第一组实验结果如图6所示，对比单传感器数据，图中红色曲线为本文算法所得，可以清晰看到本文算法能够实现多传感器信息融合，虽然是全新的一种方法，但是整体的效果还是明显的，可以应对时间域上数据源异步的问题，也就意味着可以实际使用。



(a) 全局聚类结果

(a) Global clustering results



(b) MSIF处理结果

(b) MSIF processing results

图6 算法的合理性测试

Fig.6 Algorithm reasonableness test

# 4 总结

多传感器是智能检测和自动化装置中重要的部件组成，为了解决多传感器异步数据下的融合难题，我们提出了多源异步下一种创新的基于滑动聚类的多传感器信息融合方法，核心点是我们引入了K-Means聚类方法，主要通过对k取值进行自适应设计替代人为指定，创新地设计了自适应参数的聚类滤波核，形成在时空域上滑动聚合的多传感器信息融合方法。仿真实验验证了本文方法的正确性以及合理性，可以应对时间域上数据源异步的问题。

# 参考文献：

1. Enrico S. Multimodal Sensor Fusion with Object Detection Networks for Automated Driving [D]. Chemnitz University of Technology, Germany, 2022.
2. Salman Q, Muhammad K D, Furqan Q S, et al. Multisource Data Fusion Framework for Land Use/Land Cover Classification Using Machine Vision[J]. Journal of Sensors, 2017, 2017:1-8.
3. C. Yang, Q. Chen, Y. Yang, et.al. A Deep Learning Based Highly Accurate SLAM Using Monocular Camera Aiming at Indoor Map Reconstruction with Semantic and Depth Fusion, IEEE Access, 2022(10):10259-10272.
4. 杨明月,寿莹鑫,唐勇,刘畅,许斌.多四旋翼无人机编队保持与避碰控制[J].航空学报,2022,(03):1-11.
5. 蒋筱朵,赵晓琛,冒添逸,等人.采用传感器融合网络的单光子激光雷达成像方法[J].红外与激光工程,2022,51(02):49-55.
6. 黄衍标,罗广岳,何铭金.BP神经网络在巡逻机器人多传感器数据融合中的应用[J].传感技术学报,2016,29(12):1936-1940.
7. Meng T，Jing X，Yan Z，et al. A survey on machine learning for data fusion[J]. Information Fusion,2019,57:115-129.
8. 林水生,卫伯言,杨海芬,等人,于良.引入新数据源的D-S融合检测方法[J].电子科技大学学报,2021,50(06):861-867.
9. Pan D, Liu H, Qu D, et al. Human Falling Detection Algorithm Based on Multisensor Data Fusion with SVM[J]. Mobile Information Systems, 2020, 2020(7):1-9.
10. Gong W F, Chen H, Zhang Z H, et al. A Novel Deep Learning Method for Intelligent Fault Diagnosis of Rotating Machinery Based on Improved CNN-SVM and Multichannel Data Fusion[J]. Sensors, 2019, 19(7).
11. 杨菊花,李文元,陈光武,等人.引入滑模观测器的GPS/INS组合导航滤波方法[J].仪器仪表学报,2019,40(09):78-86.
12. Jondhale S R, Deshpande R S. "Kalman Filtering Framework-Based Real Time Target Tracking in Wireless Sensor Networks Using Generalized Regression Neural Networks[J]. IEEE Sensors Journal, 2019,19(1):224-233.
13. Khamseh H B, Ghorbani S, Janabi-Sharifi F. Unscented Kalman Filter State Estimation for Manipulating Unmanned Aerial Vehicles[J]. Aerosp. Sci. Technol. 2019, 92, 446–463.
14. Liu B, Zhang T, Li Y, et al. Kernel Probabilistic K-Means Clustering[J]. Sensors, 2021, 21(5):1892.
15. Esfandiari, H., Mirrokni, V., & Zhong, P. Almost Linear Time Density Level Set Estimation via DBSCAN[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021:7349-7357.
16. Hn A, Xn B. Agglomerative over segmentation using dual similarity and entropy rate[J]. Pattern Recognition, 2019, 93:324-336.
17. Scitovski R, Sabo K. A combination of k-means and DBSCAN algorithm for solving the multiple generalized circle detection problem[J]. Advances in Data Analysis and Classification, 2021, 15:83-98.
18. 刘凤.K-means算法类簇中心优化研究[D].贵州大学, 2021.
19. Liu B, Zhang, T, Li Y, et al. Kernel Probabilistic K-Means Clustering[J]. Sensors,2021, 21:1892
20. 纪德洋,金锋,冬雷,等人.基于皮尔逊相关系数的光伏电站数据修复[J].中国电机工程学报,2022,42(04):1514-1523.
21. 高德民,钱焕延,徐江,王晓楠.无线传感器网络随机分布模型及覆盖控制研究[J].传感技术学报,2011,24(03):412-417.
22. 赵博文,王灵矫,郭华.基于泊松分布的加权朴素贝叶斯文本分类算法[J].计算机工程,2020,46(04):91-96.