Multi-sensor data fusion algorithm based on the improved weighting factor

Yongjin Yu¹, Xiaoqing Liu¹, Chuannuo Xu¹ and Xuezhen Cheng*¹

¹ College of Electrical Engineering and Automation, Shandong University of Science and Technology, Qingdao, Shandong, 266590, China
*Corresponding author’s e-mail: chengxuezhen@sdust.edu.cn

Abstract. Aiming at the decrease of the accuracy of fusion data caused by the abnormal value and noise interference in the multi-sensor observations, this paper proposes a multi-sensor data fusion algorithm based on improved weighting factors. Firstly, the Dixon criterion is used to eliminate outliers in observations to avoid data containing gross errors. Then the Kalman filter algorithm is used to effectively reduce the noise impact caused by various reasons and provides the optimal data for weighted data fusion. Finally, an improved weighted fusion algorithm is used to comprehensively consider the nature of the sensor and the influence of various factors in the measurement process to obtain the best fusion data. The simulation analysis of the soil humidity in the greenhouse shows that the error of the multi-sensor data fusion algorithm based on the improved weighting factor is maintained at 0.04%-0.18%. Compared with the adaptive weighted fusion algorithm, the error of this algorithm is reduced by 0.12%, which verifies the algorithm’s effectiveness.

1. Introduction

Multi-sensor data fusion is an algorithm for analyzing, sorting and fusing data obtained by sensor technology to obtain high-value information. The algorithm is widely used in military fields, robotics systems and industrial control fields[1]. It is precise because of the versatility of the algorithm and the diversification of application fields that its research has attracted people's attention.

Due to the variability of the monitoring environment and the error response of the system equipment itself, the data measured by the multi-sensor has problems such as uncertainty, mainly including the generation of abnormal values and noise[2]. In response to the above problems, the literature[3] uses the membership function in the fuzzy set theory to construct a support matrix and assigns weights according to the matrix, but the weights are adjusted with each previous estimate, which leads to a large error in the first measurement. Yang et al. proposed to replace the outliers eliminated in the consistency test with the most supporting data, and uses the adaptive weighted fusion estimation algorithm to obtain the fusion result, but the amount of calculation is large[4]. A weighting algorithm based on an improved support function for data fusion was suggested, which processing to improve the accuracy of fusion data[5]. Li et al. choose a two-stage fusion method, but the algorithm does not deal with data redundancy and noise[6]. The thesis improved the D-S evidence theory algorithm, and introduced the support degree in the evidence fusion stage to modify the results of the iterative fusion of evidence in [7]. Literature[8] conducts statistical analysis on the measured data of sensors, and then determines the fusion weight, but it limits the number of sensors to not more than three, which reduces the efficiency of data fusion. Jing et al. proposed to construct weights by the product of adaptive weighted weighting coefficients and the consistent mean value of support-based data fusion, to obtain the optimal fusion result, without...
analyzing the reliability of the measured value[9].

Combined with the analysis of the above-mentioned literature, to improve the accuracy of the measured value and the fusion precision, this paper proposes a multi-sensor data fusion algorithm based on an improved weighting factor. The algorithm first uses the Dixon criterion to detect and eliminate abnormal values in the data measured by the sensor. Then the algorithm combines the Kalman filter algorithm and the improved weighted fusion algorithm to fuse the pre-processed data to obtain the best-fused data; finally, the effectiveness of the algorithm is verified with experimental data.

2. Multi-sensor data fusion algorithm

In an actual detection system, the size of the detection space directly leads to changes in the amount of data collected, and the detection environment and the sensor's performance parameters will affect the validity of the data. Given the above analysis, this paper adopts the method of combining Dixon criterion and multi-sensor data fusion algorithm. This method puts the measured values into the algorithm model, and uses certain rules to obtain more scientific and comprehensive information.

2.1. Data preprocessing

In the process of system measurement, the measurement data with gross error can anti-factor produce abnormal values, such as sensor failure or its accuracy, communication transmission error, and so on. However, these abnormal values will inevitably affect data information processing, so this article uses the Dixon criterion for data preprocessing, eliminate abnormal values, ensure the accuracy of the data[10].

The data preprocessing process is as follows:

- Group the collected data and group them for preprocessing. Arrange the grouped data in ascending order, \( x_1, x_2, \ldots, x_n \).
- Calculate the statistics \( \gamma_j \) based on the above data \( x_1, x_2, \ldots, x_n \).[11].
- Determine which values are abnormal values in the group according to statistics. Judgment rules:
  - \( \gamma_j > \gamma_j', \gamma_j > C(\alpha, n) \), \( x_{o} \) is the outlier that should be eliminated.
  - \( \gamma_j > \gamma_j', \gamma_j > C(\alpha, n) \), \( x_{o} \) is the outlier that should be eliminated[12].

2.2. Kalman filter

The pre-processed data is still mixed with measurement noise and various interference signals. To effectively control the influence of noise on data fusion, the Kalman filter algorithm is used to optimize the pre-processed data information, to get closer to the true value as optimal data set.

The Kalman filter used in this article was first proposed by R. E. Kalman in 1960, and is mainly used to solve the estimation problem in linear systems. In a linear system, the Kalman filter is the optimal filter, and the application effect of this algorithm mainly depends on the accuracy of its model establishment[13].

The system state equation:

\[
X_k = F_k X_{k-1} + B_k \mu_k + W_k
\]  

(1)

Where \( X_k \) and \( X_{k-1} \) are the state values of the system at \( k \) and \( k-1 \), respectively. And \( F_k \) is the state transition matrix, \( B_k \) is the control matrix, \( W_k \) is the process noise \( W_k \sim N(0, Q) \).

State observation equation:

\[
Z_k = H_k X_k + V_k
\]  

(2)

Where \( Z_k \) is the measurement value of the corresponding state, \( H_k \) is the measurement matrix, \( V_k \) is the observation noise, and \( V_k \sim N(0, R) \). Moreover, process noise and observation noise are independent of each other.
The predicted value at time $k$ is $\hat{x}_{k|k-1}$, as shown in Equation (3). The predicted covariance at time $k$ can be expressed as $P_{k|k-1}$, as shown in Equation (4). The optimal Kalman gain is $K_k$, as shown in Equation (5).

$$\hat{x}_{k|k-1} = F_{k} \hat{x}_{k-1|k-1} + B_k \mu_k$$ (3)

$$P_{k|k-1} = F_k P_{k-1|k-1} F_k^T + Q_k$$ (4)

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1}$$ (5)

The best estimate at time $k$ is $\hat{x}_{k|k}$, as shown in Equation (6). The simplified error covariance matrix is $P_{k|k}$, as shown in Equation (7).

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \Delta_k$$ (6)

$$P_{k|k} = (I - K_k H_k) P_{k|k-1}$$ (7)

According to the description of the above equation, it can be divided into prediction process and correction process. The prediction process is to predict the current state based on the estimated value at the previous moment; the correction process is to combine the observed value and estimated value at the current moment to obtain an optimal estimated value closer to the true value[14]. The algorithm continues to iterate according to the above process and only needs to save the calculation parameters at the previous moment, without occupying a lot of storage space, making the algorithm more efficient[15].

2.3. The improved weighted fusion algorithm

The traditional adaptive weighted fusion algorithm is based on the minimum mean square error of the measured value to calculate the optimal weight factor[16], the algorithm uses the average value of the initial measured value of the sensor as the reference value, and does not consider the accuracy of the measured data. Less than the high precision requirements of the system. Based on the above situation, the weight factor is improved based on the adaptive weighted fusion algorithm, which is divided into fixed weight factor and dynamic weight factor.

2.3.1. Selection of fixed weight factors. The selection of the fixed weight factor is based on the accuracy of the sensor. This value is a fixed value and is proportional to the fusion accuracy. Therefore, the error of the sensor is assumed to be $\varphi_i$, $\varphi = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \cdots \\ \delta_n \end{bmatrix}$. And the fixed factor is defined as $\omega_i$, as shown in Equation (8).

$$\omega_i = \sqrt{\varphi_i^2}$$, $i = 1, 2 \cdots n$ (8)

2.3.2. Selection of dynamic weight factors. The selection of dynamic weight factor is based on the variance of the optimal data after the Kalman filter, and is obtained by the method of optimal weight distribution. The value will be updated in real-time as the data information is processed, that is, the closer the data is to the true value, the larger the weight factor, and the higher the fusion accuracy. The main contents are as follows:

Assuming that the $n$ optimal values in the optimal data set are $X_1, X_2 \cdots X_n$, there are unbiased estimates of the true value $X$, and their variances are $\sigma_1^2, \sigma_2^2 \cdots \sigma_n^2$, then the corresponding weight factors are $\omega_1, \omega_2 \cdots \omega_n$, and the final fusion result is $X_\mu$, as shown in Equation (9). And according to the minimum mean square error $M_\mu$, as shown in Equation (10).

$$\hat{X}_\mu = \frac{1}{\sigma_1^2 + \cdots + \sigma_n^2} \left( \frac{M_1 X_1}{\sigma_1^2} + \cdots + \frac{M_n X_n}{\sigma_n^2} \right)$$ (9)

$$M_\mu = \frac{1}{\sigma_1^2 + \cdots + \sigma_n^2}$$ (10)
The average error of the optimal data set is $\bar{X}$, then its variance is $\sigma^2$, as shown in Equation (11).

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{N} \sum_{j=1}^{N} (X_{ij} - \bar{X}) - (X_{ij} - \bar{X}) \right)^2$$  \hspace{1cm} (11)

The dynamic weight factor is $\omega_{ix}$, as shown in Equation (12). And the best fusion result is $X_i^e$, as shown in Equation (13).

$$\omega_{ix} = \frac{1}{\sigma_{i}^2} \sum_{i=1}^{N} \frac{1}{\sigma_{ix}^2}$$ \hspace{1cm} (12)

$$X_i^e = \sum_{i=1}^{N} \left[ \left( \omega_{ix} + \omega_{ix} \right) X_i \right]$$ \hspace{1cm} (13)

3. Simulation analysis of experimental data

Taking the humidity in the greenhouse as an example, the humidity data measured by the sensor is used to verify the effectiveness of the algorithm. This experimental case uses five sensors to monitor the humidity data in one day. Figure 1 is the humidity data monitored by sensor 1 in one day. According to the figure, the measured value curve fluctuates greatly and there are many abnormal data. Figure 2 is the curve after data preprocessing. The curve in Figure 2 tends to be stable, which proves that the Dixon criterion effectively removes the outliers with gross errors and enhances the anti-interference of the system, thus verifying the feasibility of data preprocessing is improved. Figures 3 and 4 are respectively the error analysis of the adaptive weighting algorithm and the improved weighted fusion algorithm, comparing the final fusion result with the error of the optimal data set.
4. Conclusion
Due to the uncertainty of multi-sensor observations, this paper proposes a multi-sensor data fusion algorithm with improved weight factors. First, the optimal data set is obtained after the two-step data processing of the Dixon criterion and Kalman filter, which effectively avoids the uncertainty of the original observation value. Then, the improved weight factor is obtained according to the dynamic weight and the solid weight, and then the best fusion data is obtained. Finally, a comparative analysis with the adaptive weighted fusion algorithm verifies the feasibility of the algorithm.

Acknowledgments
Authors wishing to acknowledge assistance or encouragement from supervisor and classmates in the laboratory. This study was supported by the Major Research Development Program of Shandong Province of China (No. 2016GSF117009), and the National Natural Science Foundation of China Program (No.62073198).

References
[1] Luo, X. Chang, X. (2015) A novel data fusion scheme using grey model and extreme learning machine in wireless sensor networks. International Journal of Control, 13: 539-546.
[2] Zhang, P. Dong, W. H. Gao, D. D. (2014) An optimal method of data fusion for multi-sensors based on Bayesian estimation. Chinese Journal of Sensors and Actuators, 27: 643-648.
[3] Yang, J. Gong, F. X. (2011) Improved dynamic weighted multi-sensors data fusion algorithm. Computer Engineering, 11: 97-99.
[4] Ding, H. Zhong, Y. Zhang, J. et al. (2016) Multi-Sensor adaptive weighted fusion algorithm based on correlation function. Journal of Chongqing University of Technology (Natural Science), 30: 114-118.
[5] Duan, Q. L. Xiao, X. Y. Liu, Y. R. et al. (2017) Data fusion method of livestock and poultry breeding internet of things based on improved support function. Transactions of the Chinese Society of Agricultural Engineering, 33: 239-245.
[6] Li, F. (2019) Application research of multi-sensor data fusion technology in greenhouse control system. Science & Technology Vision, 30: 34-36.
[7] Tian, M. M. Ye, J. H. Wang, S. M. et al. (2019) A method of multi-sensor data fusion under the complicated environment. Journal of Shandong University (Engineering Science), 49: 32-38.
[8] Yang, J. J. Ye, C. L. Feng, S. H. et al. (2020) Method of data weighted fusion for multi-sensors. Transducer and Microsystem Technologies, 39: 39-42.
[9] Jing, R. X. Gao, Y. Z. (2020) Research on data fusion algorithm based on multi-sensor. Modern Electronics Technique, 43: 10-13.
[10] Wang, Z. Bai, X. Z. Ma, M. B. et al. (2017) A data fusion algorithm on data preprocessing and Kalman filter for greenhouse environment monitor. Chinese Journal of Sensors and Actuators, 30: 1525-1530.
[11] Chen, H. Ouyang, B. Sun, L. J. et al. (2020) Application research on data fusion and fuzzy control method in greenhouse. Journal of Zhejiang University of Science and Technology, 32: 197-202.
[12] Sun, L. F. Zhang, Y. Y. Zheng, G. Q. et al. (2018) Approach to decision fusion for intelligent greenhouse environmental control based on D-S evidence theory. Journal of Agricultural Machinery, 49: 268-275.
[13] Zhou, Y. Q. Xue, H. R. Jiang, X. H. et al. (2018) Meteorological data fusion based on proposed Kalman filter method. Computer Systems & Applications, 27: 184-189.
[14] Duan, J. Jiang, Y. Tang, Y. W. et al. (2018) Research on data fusion processing technology of agricultural greenhouse based on Kalman filtering algorithm. Journal of Chinese Agricultural Mechanization, 39: 60-63.
[15] Liu, M. T. Tian, Z. Z. Qi, H. Q. et al. (2016) Sediment concentration measurement based on
Kalman-BP collaborative fusion model. Journal of Basic Science and Engineering, 24: 970-977.

[16] Yang, X. D. Wang, Y. F. Zhang, X. Q. (2017) Dynamic weighting fusion based on Kalman filter. Journal of Sichuan University (Natural Science Edition), 54: 947-952.