Multi-sensor Data Processing and Fusing Based on Kalman Filtering

Bian Guangrong*, Li Hongsheng, He ninghui
Higher Educational Key Laboratory for Automation of Hubei Province
Wuhan University of Technology, Hubei Province, China
Air Force Service College, Jiangsu Province, China
*Corresponding author, e-mail: 44973901@qq.com

Abstract
The background of this paper is the warehouse target localization and tracking system which is composed of a number of wireless sensor nodes. Firstly this paper established a model of warehouse target localization and tracking system, then a model of multi-sensor data preprocessing and data fusion was established, and self-adaptive linear recursive method was used to eliminate outliers of the original measured data. Then least squares fitting filter was used to do filtering and denoising for the measured data. In the end, the data which were measured by multi-sensor can be fused by Kalman Filtering algorithm. Data simulation analysis shows that the use of kalman filtering algorithm for the fusion of the data measured by multi-sensor is to obtain more accurate warehouse target location data, so as to increase the positioning and tracking accuracy of the warehouse target localization and tracking system.

Key Words: Wireless Sensor Network, Data Fusion, Kalman Filtering

1. Introduction
As a new technology to acquire and process information, wireless sensor networks (WSN) get serious concern in the world. In wireless sensor networks with large scale, the data fusion technology must be used before each sensor node sends the data. It collected to the convergent nodes to remove redundant invalid information and reduce the amount of data transmission, so as to reduce energy consumption, and achieve the goal to prolonging network survival period. Data fusion which is also called information fusion is a kind of multi-source information processing technology. Data fusion system is an information processing system with the purpose to improve information measuring accuracy, reliability and redundancy of the sensors [1].

The object that multi-sensor data fusion study is all kinds of information the sensors collected which are signified with the form of signal, waveform, image, data, word, sound and so on. Multi-sensor data fusion is about the various, multi-level process of the multi-sensor data to obtain more reliable and accurate data information. Multi-sensor data fusion can make comprehensive treatment to the information provided by several or several kinds of sensors to obtain useful information and do effective combination to get new information.

Multi-sensor data fusion has various methods, such as integrated average, Kalman Filtering (KF), Bayesian estimation, Dempster-Shafer evidence theory, fuzzy logic and neural network and so on. Kalman filtering is applied to fusing the real-time dynamic multi-sensor redundant data from the lower layers. This method estimates the fusing data recursively by using the statistical properties of the measurement model. This paper first established a model of warehouse target localization and tracking. Then self-adaptive linear recursive method was used to eliminate outliers in the original measured data. As the least square method is used to filter noise of the data collected by sensor nodes, Kalman filtering algorithm is used for the fusion of the data which is collected by multi-sensor.
2. Background
2.1. Problem Statement

Warehouse target localization and tracking system is composed of several wireless sensor nodes. These nodes contain beacon nodes and unknown nodes. The proportion of beacon nodes in the network is small. Beacon nodes can get their own precise location by some means such as carrying GPS positioning equipments. Beacon nodes are the reference points of unknown nodes location. And beacon nodes are arranged inside and outside the warehouse evenly. The unknown nodes can be active nodes of personnel, vehicles and warehouse equipments. By communicating with nearby beacon nodes or the unknown nodes which have acquired their own position information, the unknown nodes can calculate their own position according to a certain location algorithm. As is shown in Figure 1.

The warehouse target localization and tracking system is obviously composed of a large number of sensor nodes. The data information acquired by these sensor nodes includes a lot of redundant invalid information. So it is essential to process and fuse data by multi-sensor data fusion technology. The data fusion of multi-sensor can be described as follows. Firstly, we upload the data observed by multiple sensors in different geographical position to the data processing and fusion center. After steps of eliminating the outliers, filtering, time synchronization, data fusion and so on, new effective combined information can be acquired. The new information after combination can make full use of the complementarity between sensors, and can fully and accurately reflect the characteristics of the environment, and the result is better than the information of single sensor [1] [2].

2.2. The Model of Warehouse Target Localization and Tracking

The basic thought of wireless sensor networks localization can be represented as follows. Some special nodes are deployed in the wireless sensor networks. This kind of nodes which are called beacon nodes also have strong energy and can be equipped with GPS positioning system, or can acquire their own coordinates by other ways. By measuring the distance and angle between unknown nodes and beacon nodes or doing certain calculation according to the relative position relationship, their own coordinates can be worked out [6].

The positioning principle of warehouse target localization and tracking system is to calculate the coordinates of the unknown nodes with node position calculation method. The maximum likelihood estimation method of the multilateral measurement is used to calculate the coordinates of the unknown nodes [4].

The multilateral measurement method is often used in solving the coordinates of the unknown nodes. As shown in Figure 2, there are n reference nodes $Q_1(x_1,y_1), Q_2(x_2,y_2),...,Q_n(x_n,y_n)$, and their distance to unknown node N respectively is $r_1(t), r_2(t),..., r_n(t)$. We set the coordinates of N as $(x(t), y(t))$. Then the coordinates satisfy the equation (1).

![Figure 1. Schematic diagram of warehouse target localization and tracking system](image)

![Figure 2. Maximum likelihood estimation of multilateral measurement method](image)
The equation (2) can be acquired by the maximum likelihood estimation method.

\[
\begin{align*}
(x(t) - x_n)^2 + (y(t) - y_n)^2 &= r_n^2(t) \\
& \quad \ldots \ldots \\
(x(t) - x_{n-1})^2 + (y(t) - y_{n-1})^2 &= r_{n-1}^2(t) \\
\end{align*}
\]

(1)

\[
\begin{align*}
x_n^2 - x_n^2 - 2(x_n - x_n)x + y_n^2 - y_n^2 \\
-2(y_n - y_n)y &= r_n^2(t) - r_n^2(t) \\
& \quad \ldots \ldots \\
x_{n-1}^2 - x_{n-1}^2 - 2(x_{n-1} - x_n)x + y_{n-1}^2 - y_{n-1}^2 \\
-2(y_{n-1} - y_n)y &= r_{n-1}^2(t) - r_{n-1}^2(t) \\
\end{align*}
\]

(2)

Using system of linear equations, it can be expressed as \( AX(t) = b(t) \), where

\[
A = \begin{bmatrix}
2(x_1 - x_n) & 2(y_1 - y_n) \\
2(x_2 - x_n) & 2(y_2 - y_n) \\
\vdots & \vdots \\
2(x_{n-1} - x_n) & 2(y_{n-1} - y_n)
\end{bmatrix},
X(t) = \begin{bmatrix}
x(t) \\
y(t)
\end{bmatrix}
\]

\[
b(t) = \begin{bmatrix}
x_1^2 - 2x_1y_1 + y_1^2 - y_n^2 + r_1^2(t) - r_n^2(t) \\
\vdots \\
x_{n-1}^2 - 2x_{n-1}y_{n-1} + y_{n-1}^2 - y_n^2 + r_{n-1}^2(t) - r_n^2(t)
\end{bmatrix}
\]

The coordinates of the node \( N \) can be acquired by standard minimum mean variance estimation method.

The coordinates is

\[
\hat{X}(t) = (A^T A)^{-1} A^T b(t)
\]

(3)

3. Establishing the Model of Multi-Sensor Data Processing and Data Fusion

The process model of multi-sensor data processing and data fusion is shown in Figure 3. Firstly, sensor information acquisition system converts the target information collected by multiple sensors to digital signals, and save them in the FIFO momentarily. Then data preprocessing system preprocesses the digital signal collected by sensors. Namely, process such as eliminating the outliers, filtering and denoising and time synchronization and so on, and then measuring signals with higher quality will be got. Then data fusion system does data fusion for the measuring signals after preprocessed to output the fusion results. [10]

3.1. Sensor Information Acquisition System

The main function of sensor information acquisition system is to collect the information of measured target and uploading the collected data information to upper computers via Ethernet. As shown in Figure 4, sensor information acquisition system is composed of acquisition parts and transmission parts. The collected signals generally include analog signal and digital signal. Sensor transfers the collected data information to the controller, then the controller store the data in FIFO (First Input First Output) temporarily. Micro Controller MCU (Micro Controller with Ultra low power) inquires whether there are any data in FIFO1 and FIFO2. If there are some data, the data will be read and upload to upper computer via Ethernet.
3.2. Data Preprocessing

Warehouse target localization and tracking system is equipped with multiple sensors. These sensors may be different kinds of sensors, and their measuring principles are also different. The data collected by different sensors are also different in format and quantity. The crystal oscillator frequency of different node exist deviation. Humidity and electromagnetic interference and so on also will cause the deviation of the running time between network nodes. So, Data preprocessing for the data of the sensors before data fusion is very necessary which can ensure the quality of the data measured by sensors, and improve the effect of data fusion. Data preprocessing includes eliminating the outliers, filtering and time synchronization, etc. [1] [2]

![Figure 3. Multi-sensor data processing and data fusion process model diagram](image)

3.2.1. Eliminate Outliers

When multi-sensor collects the data, because of being interfered by atmospheric temperature and humidity, air pressure and electromagnetism, it will appear several abnormal values which are called as outliers. Outliers will have negative effects on data fusion. So we must eliminate outliers during the process of collecting the data information. Recently, there are many researches about eliminating outliers. The main methods are resistance to outliers being based on M estimation, improved $3\sigma$ eliminating outliers and adaptive resistance to outliers. The self-adaptive linear recursive method was used to eliminate outliers in this paper. [1]

Firstly, we use least square method to fit the data $(x_i, y_i)$ which are observed by sensors. We can get a group of polynomial coefficient $b_0, b_1, \ldots b_n$. Then we recursively calculate out a group of polynomial data by using the polynomial coefficient we get and the formula (4).

$$x_b(j) = \sum_{i=0}^{n} b_j j^i, j = 1, \ldots, N$$  \hspace{1cm} (4)

In the formula (4), $b_j$ is the coefficient of the fitting polynomial, $n$ is the order of the polynomial, $i$ is the Error! Hyperlink reference not valid. frequency, $j$ is the Error! Hyperlink reference not valid. time, $N$ is the data number of the least square method fitting, which can be selected as needed.

Then calculate the data’s rate of change minutely under this sample time. The specific calculation method as follow: [1]
In the formula, $r_j$ is the data’s rate of change, $x(j)$ is measurement data, $x(j)_h$ is the data which correspond with the fitting curve. Summate and average the N data’s rate of change that are got from the formula. The arithmetic as follow:

$$
\bar{r} = \frac{\sum_{j=1}^{N} r(j)}{N}
$$

(5)

After getting the average rate of change of data, refresh the data set which is used for fitting as soon as receiving data, then update the fitting curve continually, and define a range of judging outliers ($-L|\bar{r}|, L|\bar{r}|$). Calculate average of the first N historical data’s rate of change, then multiply this average by a coefficient to get a value. As this value is regard to the biggest range of current data’s rate of change, it can be the basis of judging outliers. When the absolute value of a data change rate is greater than this range, we regard the data as an outlier. After finding an outlier, use linear recursive method to replace the outlier point. The replaced point could be calculated out by formula (6).

$$
y_n = a_1 y_{n-1} + a_2 y_{n-2} + a_3 y_{n-3} + \cdots + a_k y_{n-k} (n \geq k)
$$

(6)

In the formula, $a_1, a_2, a_3, \ldots, a_k$ are constants, which are decided as needed.

If more than three outliers appear continuously, we can only use an observed value as the next value. Through this method it can well realize eliminating outliers from the measurement data and replace it. Select a group of measurement data of the warehouse target with white noises and outliers, and use the above method of eliminating outliers to simulate. The results of simulation are shown in Figure 5 and Figure 6. Figure 5 describes the original data with noises and outliers, and Figure 6 describes the original data with noises after eliminating outliers.

3.2.2. Filtering

The data information measured by sensors often is mingled with random disturbance signal. In order to reduce the influence of interference, it is necessary to smooth the collected data. The existing filtering method includes kalman filtering, low pass filtering, mean filtering, median filtering, etc. In this paper the least squares curve fitting method was adopted for data filtering processing. The basic principle of least squares filtering can be described as follows. Look for a function $S(x)$ to represent the change trend of measured data approximately. And
make the sum of squares of the error of function value and the measured data minimum at each moment.

The measured data of given sensor is \( (x_i, y_i) \) \((i=1,2,...,n)\), we set the form of fitting function is \( S(x) = a_0 \phi_0(x) + a_1 \phi_1(x) + \cdots + a_n \phi_n(x) \), where \( \phi_i(x) (k = 0,1,...,n) \) is the known linearly independent function. Now, the coefficient of function \( S(x) = a_0, a_1, \cdots, a_n \), should be worked out to make
\[
\sum_{i=1}^{n} [S(x_i) - y_i]^2 = \sum_{i=1}^{n} \left[ \sum_{k=0}^{m} a_k \phi_k(x_i) - y_i \right]^2
\]
be minimum, if
\[
\sum_{i=1}^{n} \left[ \sum_{k=0}^{m} a_k \phi_k(x_i) - y_i \right]^2 = \min_{S(x) \in \mathcal{S}} \sum_{i=1}^{n} \left[ \sum_{k=0}^{m} a_k \phi_k(x_i) - y_i \right]^2
\]
then we call the homologous
\[
S(x) = a_0^* \phi_0(x) + a_1^* \phi_1(x) + \cdots + a_n^* \phi_n(x)
\]
as least squares fitting function.

Especially, if
\[
S(x) = a_0^* + a_1^* x + \cdots + a_n^* x^m
\]
We name \( S(x) \) with \( n \) times least squares fitting polynomial. The simulation results filtering by using least squares curve fitting method are shown in Figure 7.

Figure 7. The data of the warehouse target's location after filtering

### 3.2.3. Time Synchronization

In distributed wireless sensor networks, each sensor node has its own local clock. There is a deviation of the different nodes of the crystal oscillator frequency, humidity, and the electromagnetic interference can also cause deviation of the running time between the network nodes. In the warehouse target location and tracking system, sensor node records the location and time of the target, and sends it to the gateway aggregation node, and then combines these information to estimate the position and velocity of the target. If the sensor nodes lack of uniform time synchronization, the target position is estimated to be inaccurate [7]. This paper intends to adopt TPSN time synchronization mechanism to achieve time synchronization of all sensor nodes in the entire wireless sensor networks [3].

In TPSN time synchronization protocol each sensor node is assumed to have a unique identification number ID. The wireless communication link between the nodes is a two-way channel, through two-way information exchange between the nodes it can realize the time synchronization. TPSN agreement will manage all nodes in the network according to the hierarchical structure, and is responsible for the formation and maintenance of hierarchical structure [4][5].
(1) TPSN agreement is divided into two stages
The first stage: creating hierarchical structure. Each node is endowed with a level. The root node are endowed with the highest level of level 0 and i-level node can at least communicate with one of the (i-1)-level nodes. The second stage: all the tree nodes realize time synchronization. 1-level node is synchronous to the root node, i-level node is synchronous to (i-1)-level node, in the end all nodes are synchronous to the root node and realize the network time synchronization. [8]

(2) The synchronous mechanism between adjacent level nodes
The model of synchronism between adjacent level nodes is shown as Figure 8. Node i send synchronous messages (SYNC_MSG) to node i-1 in T1 moment, node i-1 record arrival time T2 of synchronous message with own clock, T2 τ = T1 + D + δ, among them, D is message transmission time. δ is the clock offset between nodes. Then, in T3 moment i-1 node return a acknowledgement message (ACK_MSG) including T2 and T3 to node i, the i node received this acknowledgement message in T4 moment, T4 = T3 + D - δ. If the send synchronous message and acknowledgement message in the process have the same transmission time, we can calculate clock offset and transmission time with the formula (9). [9]

\[ \delta = \frac{[(T_2 - T_1) - (T_4 - T_3)]}{2}, \quad D = \frac{[(T_2 - T_1) + (T_4 - T_3)]}{2} \] (7)

3.3. Data Fusion System
After pretreated, the data of multi-sensor needs to be fused in order to ensure that the final fusion data is accurate and integrated. There are many data fusion method, such as kalman filter, fuzzy neural network, wavelet neural network. In this paper, we will apply kalman filter to multi-sensor data fusion.

3.3.1. Kalman Filter Model
(1) The mathematical description of the kalman filtering
In the kalman filter, the mathematical description of the system state equation and measurement equation is as follows:

\[
x_{t+1} = A_t x_t + B_t u_t + W_t
\]

\[
y_t = C_t x_t + V_t
\] (8)

Where x is state vector, u is control vector, y is observation vector, W is process noise vector, V is the observation noise vector.

Here we define \(\hat{x}_t^-\) as \(x_t\) priori state estimate, \(\hat{x}_t^+\) as \(x_t\) posteriori state estimate value, \(\hat{y}_t^-\) as \(y_t\) priori state estimate value.

While priori estimate of the current value of a random variable is based on previous moment and earlier historical observation information, posterior estimate is based on the present moment and earlier historical observation information.

We can get the priori estimate value of \(x_t\) by the posterior estimated value of previous time and the input information. A priori estimates of \(y_t\) also could be obtained by priori estimate prediction of \(x_t\).

\[
\hat{x}_t^- = A_t \hat{x}_{t-1} + B_t u_{t-1}
\]

\[
\hat{y}_t^- = C_t \hat{x}_t^-
\] (9)

A difference between the actual measured value and predicted value of \(y_t\) is called the residual of the filtering process.
Residual = \( y_i - \hat{y}_i = y_i - C_i \hat{x}_i \) \( (10) \)

Residual reflects the difference between the forecast value and the actual value. If residual is zero, its estimated value and actual value were perfect. If the residual is small, estimate is very good. If not, it is not good. Kalman filter can use the remnants information to improve estimates of \( x_i \) and give a posteriori estimation:

\[
\hat{x}_i = \hat{x}_i^* + k_{\text{Residual}} = \hat{x}_i^* + k(y_i - C_i \hat{x}_i^*) = (1 - kC_i) \hat{x}_i^* + ky_i \tag{11}
\]

Here the \( k \) is called kalman gain.

Now the rest of the problem is how to find a suitable \( k \), to make estimation the best.

This need to define a priori error \( e_i^- \) and posterior error \( e_i^+ \):

\[
e_i^- = x_i - \hat{x}_i^* \quad e_i^+ = x_i - \hat{x}_i^* \tag{12}
\]

Their variance is the priori mean variance \( p_i^- \) and the posterior mean variance \( p_i^+ \):

\[
p_i^- = \text{var}(e_i^-) = \left\{ (e_i^-)^2 \right\} \tag{13}
\]

\[
p_i^+ = \text{var}(e_i^+) = \left\{ (e_i^+)^2 \right\} \tag{14}
\]

The best \( k \) value can make posterior mean variance to be minimum, and is the value which can make \( \frac{\partial p_i^+}{\partial k_i} = 0 \). Here \( k \) value is changing along with time, so \( k \) is replaced by \( k_i \).

According to the formula, you can calculate a priori mean variance and a posteriori mean variance:

\[
p_i^- = A_i^2 p_{i-1}^- + \left\{ W_i^2 \right\} \tag{2}
\]

\[
p_i^+ = (1 - C_i k_i) p_i^- \tag{11}
\]

Where \( k_i = C_i p_i^- / \left[ C_i^2 p_i^- + \left\{ V_i^2 \right\} \right] \]

(2) The steps of kalman filter algorithm

Kalman filter algorithm is divided into two iterative steps: time update and measurement update. This can be summed up as shown in Figure 9.

Time's update and forecast

- Predict the state of the system
  \[
  \hat{x}_i = A_i \hat{x}_{i-1} + B_i u_{i-1}
  \]

- Predict the mean variance of the system
  \[
  p_i^- = A_i^2 p_{i-1}^- + \left\{ W_i^2 \right\}
  \]

Measurement update correction

- Correction of kalman mixed coefficient
  \[
  k_i = C_i p_i^- / \left[ C_i^2 p_i^- + \left\{ V_i^2 \right\} \right]
  \]

- Correct system state
  \[
  \hat{x}_i = (1 - kC_i) \hat{x}_i^* + ky_i
  \]

- Correct the mean variance of system
  \[
  p_i^+ = (1 - C_i k_i) p_i^- \]
3.3.2. Multi-Sensor Data Fusion

Background of multi-sensor data fusion model is based on the warehouse target tracking system and wireless sensor networks. The node localization algorithm of the warehouse target tracking system is maximum likelihood estimation of multilateral measurement method. We can get the following formula by the maximum likelihood estimation of multilateral measurement method:

\[ AX(t) = b(t) \]

Where

\[ A = \begin{bmatrix} 2(x_1 - x_n) & 2(y_1 - y_n) \\ \vdots & \vdots \\ 2(x_{n-1} - x_n) & 2(y_{n-1} - y_n) \end{bmatrix}, \quad X(t) = \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} \]

\[ b(t) = \begin{bmatrix} x_1^2 - x_n^2 + y_1^2 - y_n^2 + r_1^2(t) - r_n^2(t) \\ \vdots \\ x_{n-1}^2 - x_n^2 + y_{n-1}^2 - y_n^2 + r_{n-1}^2(t) - r_n^2(t) \end{bmatrix} \]

By the above analysis, \( r_1(t), r_2(t), \ldots, r_n(t) \) is measurable. Therefore, the state equation and the measurement equation of the model is as follows.

\[ x(i, t) = Ax(i, t-1) + W \]
\[ r(i, t) = Cx(i, t) + V \]

\( x(i, t) \) —represents the state of sensor \( i \) at time \( t \);
\( r(i, t) \) —represents the measurement value of sensor \( i \) at time \( t \).

After establishing the state equation and measurement equation, according to the kalman filtering algorithm steps we can constantly update measurement and time, until the kalman gain tend to be a specific value, it make \( \hat{x}(i, t) \) optimum [10].
4. Results and Analysis

According to the model’s actual situation which is combined with warehouse target location and tracking system, assuming that there are three sensor, respectively sensor 1, sensor 2, and the sensor 3, sampling period is \( T = 1 \) s. Using kalman filtering algorithm of data fusion, the simulation of matlab graphics are shown in Figure 10, Figure 11 and Figure 12.

Figure 10, 11 and Figure 12 respectively describe three displacement sensors in the case after KF. Each figure is consist of four small figures, in which the left above diagram describe the situation of the real value (square bar head), a priori estimate (hollow circular rod head) and posteriori estimation (solid circular rod head). Obviously, a posteriori estimates approximate the true value at a very fast speed. The upper right picture describes the trend of priori variance and posteriori variance. Apparently, they approach an asymptotic value at a very fast speed. The lower left diagram depicts a priori and a posteriori error changes. Clearly, they are decreasing at a very fast speed. The lower right describes the change trend of kalman mixed coefficient. As you can see, kalman mixed coefficient approach a nonzero constant at a very fast speed.

![Figure 10. The kalman filtering schematic diagram of the second sensor](image1)

![Figure 11. The kalman filtering schematic diagram of the second sensor](image2)

![Figure 12. The kalman filtering schematic diagram of the third sensor](image3)
Figure 13 describes the relative errors of three sensors’ locations before kalman filtering data fusions. Figure 14 is the relative errors of three sensors’ locations after kalman filtering data fusions. Evidently, after kalman filter data fusion, the displacement error of sensor measurement has been obviously improved.

Figure 13. Relative errors of three Sensors’ locations before KF data fusions

Figure 14. Relative errors of three sensors’ locations after KF data fusions
5. Conclusions and Suggestions
This paper firstly established warehouse target location and tracking system model which is basing on multilateral measurement method of maximum likelihood estimation algorithm, and then established the multi-sensor data preprocessing and data fusion model. Outliers can be removed from the original measurement data by adaptive linear recursive detection method. Using the method of least squares fitting filtering to filter and denoise the measurement data, finally, the data which is measured by multi-sensor is fused with Kalman filter algorithm. The simulation results show that, outliers from the original measurement data can be effectively removed by adaptive linear recursive detection method, and using the method of least squares fitting filtering realized filtering and denoising well. Accurate warehouse target location data was obtained with multi-sensor data fusion by kalman filtering algorithm. By this way, it increase the accuracy of positioning and tracking of the warehouse target location and tracking system.

References