A LoRa-based Remote Gesture Monitoring System Using Deep Learning

Junwei Xie\textsuperscript{1,2,*}, Wei Song\textsuperscript{1}, Amanda Gozho\textsuperscript{1} and Fan Yu\textsuperscript{2}

\textsuperscript{1} School of Information Science and Technology, North China University of Technology University, Beijing, China
\textsuperscript{2} School of Geomatics and Urban Spatial Informatics, Beijing University of Civil Engineering and Architecture, China

*Corresponding author e-mail: 2108160320012@stu.bucea.edu.cn

Abstract. To solve the problems of high power consumption, low transmission distance and low recognition accuracy of the gesture monitoring system of traditional wearable devices, this paper designs a remote gesture monitoring system based on LoRa. In terms of data transmission, LoRa Internet of Things technology is used, which has the characteristics of low power consumption, high speed and long-distance transmission, and can meet the needs of multi-user long-term use. The identification module is built on the remote server and can be used directly without configuration. Based on the multi-sensor data, this paper also designs a deep learning model to complete the task of human gesture recognition, which can recognize 7 kinds of gesture data and the effect meets the expectations.

Keywords: LoRa, Internet of Things, Gesture Recognition, Wireless Sensor Network, Deep Learning

1. Introduction

Human gesture recognition is a subject of predicting human behavior, describing human gesture and studying human state. The research methods mainly include machine vision and wearable equipment [1, 2]. Wearable devices are popular in people’s daily life because of their low price, low energy consumption and high sensitivity. With the rapid development of Internet of Things technology, artificial intelligence technology and information communication technology, human gesture recognition is gradually combined with wearable devices. Wearable devices submit sensor data to the cloud server for analysis, which gradually becomes the mainstream research direction of a human gesture recognition system [3, 4].

2. System Design Scheme

The remote gesture monitoring system based on LoRa proposed in this paper is shown in Fig. 1. The user wears the hardware device at the waist, and the hardware device will collect the user’s motion data and transmit the data to the LoRa base station through the LoRa node. As an intermediary, LoRa base station sends data to remote server through Internet router. The server receives the data and
analyzes the data through deep learning, and feeds the analysis results back to the front end of monitoring system for visualization.

### Figure 1
The system is composed of four modules: information collection module, data upload module, gesture analysis module and result display module.

The information collection module is the hardware equipment that users wear. It is composed of a three-axis acceleration sensor, a two-axis gyroscope sensor and a microprocessor, which is used to collect user’s gesture information. The data upload module is composed of a LoRa node and a LoRa base station. The LoRa node receives the motion data from hardware and transmits the data to the LoRa base station through LoRa wireless transmission. When the LoRa base station obtains the data from the LoRa node, it will transmit the data to remote server through the Internet router. The gesture analysis module is located on the remote server, receives the data sent from the LoRa base station, filters the error data, classifies and stores the data according to the user number, and sends the classified data to the deep learning model for analysis to get the user gesture results. The result display module visualizes the user’s results obtained by the gesture analysis module, so that the monitor can monitor the motion gesture of each user in real time, so as to find the problematic gesture of the user in time and control the risk more effectively.

#### 2.1 Information Collection Module
The information collection module completes the collection of the user’s motion gesture information. In order to obtain more representative human gesture data, the device should be worn on the user’s waist, chest or back. The above data is representative in the task of human gesture recognition, so only the above two sensors are used to collect data. Compared with the scheme using only a three-axis acceleration sensor, the scheme using a three-axis acceleration sensor and two-axis gyroscope obtains additional pose angle data, which can comprehensively judge the gesture state and reduce the probability of a false alarm, especially in the alternating gestures of going up and down the stairs, fast sitting and lying down.

The microprocessor will periodically activate the sensor, which will hand the acquired gesture characteristic data to the LoRa node. The original data of the sensor will be affected by the external environment, the user’s irrelevant actions and so on, which will inevitably generate random noise. Before extracting gesture features, data needs to be pre-processed to offset the impact of random noise, so as to get a better performance effect in the deep learning model. Considering the need to upload data to remote servers via wireless transmission, due to the limitation of the wireless transmission rate, the uploaded data should be as little as possible. In order to solve the contradiction between less data uploaded and more gesture features remaining, the microprocessor should have the function of data pre-processing.

After considering the timeliness and complexity of the algorithm, the moving average filtering method is adopted. The moving average filtering method is a simple time series data processing method. Through moving average filtering, the original time series data of the sensor is replaced by the recursive average value, which can, not only retain the original periodic fluctuation characteristics...
of the data, but also eliminate the random noise caused by unstable factors [5]. The formula of moving average filtering is as follows:

\[ X_k = \frac{1}{n} \sum_{i=k}^{k+n} x_i \]  

(1)

In this formula, \( k \) represents the index of the starting point of filtering, and \( n \) represents the number of adjacent points used in a filtering operation. After many experiments, when the \( k \) value is 8, the moving average filtering can produce a better noise reduction effect.

2.2 Data Upload Module

Traditional wireless transmission technologies such as Bluetooth and Wi-Fi modules cannot meet the long-term requirements of long distance and low power consumption for user’s posture data transmission [6]. Therefore, this paper proposes a remote posture monitoring system based on LoRa. With the characteristics of LoRa technology, such as long-distance transmission, low power consumption, high transmission speed and fast response speed, the contradiction between transmission distance and low transmission power consumption in traditional infinite transmission technology is solved [7]. Only one LoRa base station needs to be arranged in an 8-story building to receive the data broadcast from the indoor LoRa node. In comparison with Bluetooth devices, the LoRa device has a sustainable battery life which allows it to be used for up to one year after one charge, which reduces the inconvenience of frequent charging [8].

The data upload process includes two aspects, from the LoRa node to the LoRa base station and from the LoRa base station to a remote server. After the device is powered on, a microprocessor will configure the parameters of the SX1278 chip, such as broadband, communication frequency, etc., and periodically activate it to check the buffer data. If there is data to be sent in the buffer, the microprocessor should send the data to the LoRa base station and clear the buffer, otherwise continue to wait for the data to be loaded to the buffer. Through the data broadcast function of the LoRa node, the data is converged to the LoRa base station. After receiving the data broadcast by each device, the LoRa base station sends the data to a remote server through the Internet router. The remote server receives the data, then stores and analyzes it.

2.3 Gesture Analysis Module

The data package of a minimum element contains several vectors composed of sensor values, which we call the gesture eigenvector. These vectors are sorted in chronological order. If \( X_i \) is set as the attitude eigenvector of the i-th moment, the gesture eigenvector can be expressed as:

\[ X_i = [ a_{x_i} \ a_{y_i} \ a_{z_i} \ y_i \ \lambda_i ]^T \]  

(2)
Where $a_{xi}$, $a_{yi}$ and $a_{zi}$ represent the acceleration values of X axis, Y axis and Z axis at the i-th moment respectively, and $\gamma_i$ and $\lambda_i$ represents the gyroscope inclination at the i-th moment. By combining N gesture features into matrix $A$, the expression of matrix $A$ is:

$$A = \begin{bmatrix}
X_i & X_{i+1} & \ldots & X_{n-1} & X_n
\end{bmatrix}$$

It can be seen that the size of matrix $A$, composed of N gesture features is $5 \times N$. Generally speaking, for the gesture features with close sampling time, they are closely related. For the gesture features with large sampling time distance, their correlation is weak. In the task of gesture recognition, it is not necessary to perceive the global gesture characteristics. As long as the local gesture features are integrated, the user gesture at a certain time can be obtained.

Convolutional neural networking is widely used in the field of image recognition, which is inspired by the biological structure of the mammalian visual cortex [9]. The $A$ matrix above can be regarded as a $5 \times N$ feature graph, and N gesture features can be classified by the convolutional neural network.

It can be seen that 2s can represent a gesture period by observing the numerical curves of sensors under different gestures. The $A$ matrix is used to contain all the gesture features in 2s, and the convolutional neural network is used to classify the $5 \times N$ feature graph. Since the shape and size of the input layer of the traditional convolutional neural network are fixed, it is necessary to ensure that the size of each input image is the same, so it is better to specify the input image size as $5 \times 20$. The feature graph should be scaled to a $5 \times 20$ feature graph by linear interpolation with the number of rows unchanged.

![Figure 3](image_url)

**Figure 3.** The CNN [10] architecture is shown in this figure. The input feature graph is a single channel image with a size of $5 \times 20$ pixels. After three convolutional layers and two full connection layers, seven gestures are output. The convolutional kernel size is $3 \times 3$, the step size is 1, and the input layer and $L_1$ layer use the edge expansion with pixel 1. The size of $L_1$ layer is $8 \times 5 \times 20$ pixels, $L_2$ layer is $16 \times 5 \times 20$ pixels, $L_3$ layer is $16 \times 3 \times 18$ pixels, $F_1$ layer is 64 pixels, and output layer is 7 pixels.

In order to prevent over fitting the deep learning model, a Dropout strategy is used in training, and all activation functions use RELU function, and a cross entropy loss function is used in loss function. SoftMax regression is used for the output of the last layer. The discrete result is transformed into the form of probability distribution, and the one with the largest probability is taken as the prediction value of user gesture.

### 3. Experiment and Result Analysis

According to the principle of cross validation, the data was divided into 10 groups, and each group had 100 pieces of data. 900 pieces of data were selected as training samples each time, and the remaining 100 pieces of data were used as test samples. This process was repeated 10 times.

**Table 1.** The confusion matrix of the gesture recognition model. 1000 pieces of data on seven gestures were collected.

|        | Walking | Standing | Laying | Stand to sit | Sit to lie | Walking up/downstairs | Fall |
|--------|---------|----------|--------|--------------|-----------|-----------------------|------|
| Walking| 137     | 0        | 0      | 2            | 6         | 12                    | 0    |
| Standing| 0       | 132      | 5      | 0            | 0         | 0                     | 0    |
Laying  0  18  144  0  0  0  0
Stand to sit  0  0  0  141  4  0  4
Sit to lie  3  0  0  5  136  0  2
Walking up/downstairs  10  0  0  0  0  134  0
Fall  0  0  1  2  4  4  194

Table 2. The classification effect

| Activity                  | Precision | Recall | Support |
|---------------------------|-----------|--------|---------|
| Walking                   | 87.26%    | 91.33% | 150     |
| Standing                  | 96.35%    | 88.00% | 150     |
| Laying                    | 88.89%    | 96.00% | 150     |
| Stand to sit              | 94.63%    | 94.00% | 100     |
| Sit to lie                | 93.15%    | 90.67% | 100     |
| Walking up/downstairs     | 93.06%    | 89.33% | 150     |
| Fall                      | 94.63%    | 97.00% | 200     |

The classification effect of the gesture recognition model is shown in Table 2. According to the analysis in Table 2, the recognition rate of falling down, sitting or standing up, lying down or getting up, and going up and down the stairs are higher. The recognition rate of lying is low, and it is easy to be mistaken for standing. The average value of correct recognition is 92.57%, and the recognition rate can meet the needs of the system.

4. Conclusions
This paper designs a remote gesture monitoring system based on LoRa. In terms of data transmission, LoRa Internet of Things technology is used, which has the characteristics of low power consumption, high speed and long-distance transmission, and can meet the needs of a multi-user’s long-term use. The identification module is built on the remote server and can be used directly without configuration. Based on multi-sensor data, this paper also designs a deep learning model to complete the task of human gesture recognition, which can recognize 7 kinds of gesture data, and the effect meets the expectations. The whole system can realize the process of basic data collection, data processing and data display, however, there is still a certain gap in application to the production environment. In the future, it is necessary to improve the hardware and software functions of the system and try to improve the classification accuracy.

References
[1] Wu Y., Huang T.S. Vision-Based Gesture Recognition: A Review. In: Braffort A., Gherbi R., Gibet S., Teil D., Richardson J. (eds) Gesture-Based Communication in Human-Computer Interaction. GW 1999. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol 1739. Springer, Berlin, Heidelberg (1999)
[2] G. R. S. Murthy & R. S. Jadon.:A Review Of Vision Based Hand Gestures Recognition. International Journal of Information Technology and Knowledge Management, vol. 2(2), pp. 405-410. (2009)
[3] Chen, Chun-Wei, and Kai-Tai Song.: Multi-person pose recognition system using a zigbee wireless sensor network. U.S. Patent No. 7,742,894. (2010)
[4] Wang, Shiye, Zhezhou Yu, and Xiangchun Yu.: Real-time online action detection and segmentation using improved efficient linear search. International Journal of Computing Science and Mathematics 10.2: 129-139. (2019)
[5] Babu, C. Narendra, and B. Eswara Reddy.: A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data. Applied Soft Computing 23: 27-38. (2014)
[6] Marais, Jaco M., Reza Malekian, and Adnan M. Abu-Mahfouz.: LoRa and LoRaWAN testbeds: A review. 2017 Ieee Africon. IEEE, 2017.(2017)
[7] Vangelista, Lorenzo, Andrea Zanella, and Michele Zorzi.: Long-range IoT technologies: The
dawn of LoRa™. Future access enablers of ubiquitous and intelligent infrastructures. Springer, Cham. (2015)

[8] Becker, D.: Sensei: A Real-time Recognition, Feedback and Training System for Tai Chi Gestures. MIT Media Lab, MS thesis (1997)

[9] LeCun, Yann, et al.: Gradient-based learning applied to document recognition. Proceedings of the IEEE 86.11 (1998): 2278-2324. (1998)

[10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton.: Imagenet classification with deep convolutional neural networks.” Advances in neural information processing systems. (2012)