Gesture Recognition Control System Based on Raspberry PI

Xiangsheng Feng

School of Information Engineering, Nanchang Institute of Technology, Nanchang, Jiangxi, China, 330099
*Corresponding author e-mail: fengxiangsheng123@nit.edu.cn

Abstract. With the development of science and technology, people have been searching for a new way of man-machine interaction to break through the constraints of mouse and keyboard. Gesture recognition technology has been widely studied because of its natural and convenient interpersonal communication. As early as 1993, Thomas proposed to recognize gestures with the help of data gloves or special colored auxiliary marks pasted on the hands. Thus, people began to explore the field of gesture recognition. Through more than ten years of development, especially the rapid development of computer vision, people have gradually stepped forward from pasting special color auxiliary marks to naked hand recognition technology. Based on this, the author combined gesture control board with image recognition method to carry out two-factor authentication of gesture recognition. The proposed gesture recognition system has an accuracy rate of 90% and can be constructed by using raspberry PI as the control center.

Keywords: Raspberry Pie, Gesture Recognition, Smart Home

1. Introduction

Man-machine interface is an important link in the development of smart home. Some scholars have proposed that the hand is the most flexible part of the body, and it will vary according to each person's habits. For example, the temporal and spatial displacement of the wave can be recognized according to these characteristics. Gesture recognition has to solve the problem of how to define the beginning and end of a gesture. Background subtraction is usually used to get the start and end points of the gesture, but daily life photos can lead to false positives in some cases. Deep learning is an algorithm developed by neural networks to simulate the behavior of the human brain. Many researchers have applied it in the field of gesture recognition, and proposed to use long-term recursive convolutional network to classify gestures, with an accuracy of 91%. Some scholars proposed a multi-pattern gesture recognition method based on deep learning network[1]. The experimental results show that the PROPOSED MRNN is superior to the gesture data set method and has strong adaptability to noise interference. The above research results show that deep learning has a good recognition effect for gesture recognition, so this paper will use deep learning to conduct experiments[2].

2. Gesture recognition

In this paper, the method to determine the start of a gesture is to use the gesture sensor board to
receive signals to determine whether the gesture begins. When the gesture is confirmed to start, the camera is activated by raspberry PI to get the gesture image, while the threshold detection method is adopted to determine the end of the gesture. Threshold acquisition is to extract 10 gesture-free images at the first use, and convert the 10 images to grayscale images for matrix addition, so as to obtain the average value and obtain the end threshold of the gesture. When the camera is recording, the threshold is calculated at the same time. If the threshold of the current image is equal to the end threshold of the gesture, the end of the gesture and the end of the recording will be determined[3].

The method used for feature extraction in this paper is to cut each frame of the recorded gesture video and then obtain ten key frames by extracting key frames. Because video is composed of many frames of continuous images, extracting key frames not only reduces the data dimension, but also avoids overlearning. As shown in Figure 1, Figure 2 and Figure 3.

3. The results of four gesture experiments

The experimental data of the preliminary experiment can be divided into two parts: the training data set and the test data set. Training data set for the up and down or so every 45 group a total of 180 sets of data; Test sets up and down or so each 5 groups, each group of data contains 2560 pixels. In order
to understand whether the data types presented by different gestures are different, the training sets of four gestures are visualized to facilitate observation. We can see that the pixel and grayscale image of the four gestures have different distribution modes respectively. Therefore, it can be inferred that the feature extraction method proposed in this paper has certain representativeness for gesture recognition[4].

Based on the above conjectures, further experiments are carried out to input this training set into deep learning neural network for model training, so as to facilitate subsequent gesture recognition. This article USES the mean square error (mse) and root mean square error of the merits of the two formulas evaluation model, its value as small as possible. As shown in Equations:

\[
MSE = \frac{1}{n} \sum_{t=1}^{n} (Actual\_value_t - prediction_t)^2
\]

\[
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (Actual\_value_t - prediction_t)^2}
\]

As assessed, the MSE and RMSE values of the Tanh function are better than those of the SCR function in the training phase, but the difference is not significant, and emphasis is put on the test.

The test data set of this experiment consists of 5 sets of data at the top, the bottom, the left and the right, in total 20 sets. Use the formula to calculate the accuracy. When the predicted value is the same as the target value, it is marked as 1 and vice versa. N is the total number of groups. In the test phase, the accuracy of both activation function experiments reached 95%.

| aim   | predict | down | left  | right | up  |
|-------|---------|------|-------|-------|-----|
| 1     | up      | 0.00 | 0.00  | 0.00  | 1.00|
| 2     | up      | 1.00 | 0.00  | 0.00  | 0.00|
| 3     | up      | 0.00 | 0.00  | 0.00  | 1.00|
| 4     | up      | 0.00 | 0.00  | 0.00  | 1.00|
| 5     | up      | 0.00 | 0.00  | 0.00  | 1.00|
| 6     | down    | 1.00 | 0.00  | 0.00  | 0.00|
| 7     | down    | 1.00 | 0.00  | 0.00  | 0.00|
| 8     | down    | 1.00 | 0.00  | 0.00  | 0.00|
| 9     | down    | 1.00 | 0.00  | 0.00  | 0.00|
| 10    | down    | 1.00 | 0.00  | 0.00  | 0.00|
| 11    | left    | 0.00 | 1.00  | 0.00  | 0.00|
| 12    | left    | 0.00 | 0.99  | 0.01  | 0.00|
| 13    | left    | 0.00 | 1.00  | 0.00  | 0.00|
| 14    | left    | 0.00 | 1.00  | 0.00  | 0.00|
| 15    | left    | 0.00 | 1.00  | 0.00  | 0.00|
| 16    | right   | 0.00 | 0.99  | 0.01  | 0.00|
| 17    | right   | 0.00 | 1.00  | 0.00  | 0.00|
| 18    | right   | 0.00 | 1.00  | 0.00  | 0.00|
| 19    | right   | 0.00 | 1.00  | 0.00  | 0.00|
| 20    | right   | 0.00 | 1.00  | 0.00  | 0.00|

**Figure 4.** The relevant formula.

As can be seen from Figure 4, AIM in the table is the actual value and predict is the predicted value. Down, left, right and up are the results predicted by the deep learning neural network respectively. The error judgment of this experiment is the second set of data, and the target is upwardIt's going down. In the preliminary experiment, this paper first collects a small amount of data for testing and finds that the accuracy is 95%, indicating that the gesture recognition architecture
proposed in this paper is feasible[5].

4. Gesture segmentation in complex background
In the process of actual image acquisition, background noise may interfere with hand part extraction. The influence of background noise similar to skin color should be excluded in the design of the system. A photo frame and clothing similar in color to the skin were chosen for the background. The binarization image is obtained by segmenting the above skin tone model.

OpenCV provides a function to find connected regions, which returns the label and pixel area of each connected region. Since the hand region is dominant in the image, the connected region with the largest pixel area is chosen as the candidate region[6]. The coordinates of the connected region and the center point are marked with a rectangle. The gesture segmentation region can be obtained by filling other areas that do not meet the requirements.

The design should also consider that if the hand is not in the image capture range, the background area similar to skin color should not be mistaken for the gesture area. In order to prevent such a situation, it is stipulated that when the maximum connected area is less than 5000 pixels, it will not be marked. The system has been tested in many harsh environments and the experimental results are satisfactory. The gesture area can be correctly divided[7].

5. Conclusion
In this paper, image processing and neural network are used to realize double recognition of gesture recognition and the application of the system is realized based on raspberry PI. In the experiments of excitation function, the Tanh function performs better than SCR function in training phase. In the tests, there was little difference. Therefore, in the future, more gesture types will be combined, and noise will be added for experiments to increase the credibility of experimental results. Then, an automatic updating missing value filling model will be established for practical operation. It is also worth noting that the context is more complex, especially when the gesture areas coincide with areas similar to skin color. The system will treat the gesture part and the whole noise area as a connected area and affect the gesture recognition. How to eliminate the influence of this point and achieve better robustness will be our next step to focus on solving the problem.

Acknowledgments
Jiangxi Science and Technology Research Project of Education within the Department of China: Target tracking based on kernel affine packet representation and support vector dictionary learning (No: GJJ180939).

References
[1] Li Jing. The 41st Statistical Report on The Development of Internet in China Published by [J]. China Radio, 2018(3): 96.
[2] Deng Tianjian. Research on Human Behavior Recognition Based on Hierarchical Model [D]. Hefei:University of Science and Technology of China, 2010: 15.
[3] Zhu Sicong, Zhou Delong. Overview of Corner detection Technology [J]. Computer System Should Use, 29 (1): 22 to 28.
[4] Zhao Xuezhang, Ju Huajian, Xi Yunjiang. The combination of temporal and spatial interest points with HMM Human motion Recognition Method [J]. Application of Microcomputer, 2008, 34(12): 4-7, 13.
[5] Tao Ling. Human Behavior Recognition Algorithm based on PE-CuboID [J].Computer Knowledge and Technology, 2014(6): 133-136.
[6] Ding Chongyang. Skeleton Behavior of human body based on spatio-temporal weight attitude motion characteristics Identification research [J]. Acta Computerica Sinica, 2019(7): 1-12.
[7] Dong Nan. Study on extraction method of human body contour in static images [D]. Xi'an University of Technology, 2018:64.