Design and Performance Test of Gesture Recognition Control System Based on Machine Learning in Artificial Intelligence

Liu Tiantian

Sichuan Aerospace Vocational College, Chendu 610000, Sichuan, China

email: andyliutian@163.com

Keywords: Artificial Intelligence, Machine Learning, SVM, Gesture Recognition

Abstract: Under the background of artificial intelligence, with the arrival of big data era, the scale of data to be processed is becoming larger and larger, and machine learning has gradually penetrated into all aspects of people's life and work. With people's increasing requirements for accuracy and ease of use of gesture recovery and gesture recognition, key problems and challenges such as lack of prior knowledge representation, compromise between accuracy and delay, tedious calibration process and insufficient labeled data are superimposed alternately. In order to realize fast and accurate static gesture recognition, a new classification method based on support vector machine is proposed in this paper. By selecting proper function subset, the recognition rate of discriminant function is optimized, and a learning machine with generalization ability and optimal classification ability is obtained. The gesture recognition test is carried out by Kinect, and the results show that the gesture recognition algorithm based on improved SVM vector machine has better accuracy and accuracy.

1. Introduction

The continuous development of artificial intelligence has greatly improved the speed and level of development in the world. Among them, the development of some intelligent robots has gradually transitioned from theory to our life. Intelligent robot reduces the complexity of application through the cooperation between CPU and man-machine interface, which makes its application more and more extensive and efficient [1]. Among them, the rapid development of virtual reality technology forces people to find new ways of human-computer interaction. The development of speech recognition, gesture recognition and somatosensory technology provides a new interactive way for human-computer interaction. Both information exchange and system application are inseparable from human-computer interaction. The communication between end users and computers must be realized through various human-computer interaction interfaces [2]. Therefore, as an important part of computer science, human-computer interaction technology has attracted more and more researchers' attention.

There are various defects in the traditional man-machine interaction, such as using keyboard, mouse or joystick to complete the man-machine interaction. Therefore, new interactive ways emerged, among which dynamic gesture recognition has gradually become one of the hottest new interactive ways in human-computer interaction. In this paper, several common gesture operations are analyzed firstly, and on this basis, the machine learning method SVM is used to classify and recognize the data of 3D acceleration sensor.

2. Gesture Recognition Framework based on SVM

2.1. SVM Technology

SVM is a learning method based on statistical theory, which adopts the inductive principle of structural risk minimization. Gesture recognition mainly includes the acquisition of gesture images, the preprocessing of extracted images (including grayscale and binarization), the extraction of image geometric moments, the training of SVM models and the final matching [3]. Feature

  DOI: 10.25236/cseem.2020.126

extraction includes extracting distance information, height information and palm region features. Gesture recognition integrates all feature information into a multi-classification support vector machine to complete recognition. The principle of SVM is that points belonging to different kinds in vector space are separated by an optimal hyperplane, so that points belonging to the same kind are divided to the same side as much as possible, and the distance between the two classes is maximized. The video data stream is acquired by camera equipment, and then the system detects whether a predefined gesture appears in the data stream according to the interactive model of gesture input, and if so, the gesture is segmented from the video signal. Based on the above characteristics of SVM, it is superior to other pattern recognition methods when dealing with small sample, nonlinear and high dimensional data.

2.2. Gesture Recognition Framework Based on SVM

The decision function of support vector machine algorithm is determined by a few support vectors calculated from the training samples, and the complexity of the algorithm depends on the number of support vectors, not on the spatial dimension of the samples [4]. It can not only extract features from single-channel images, but also extract features from each channel in multi-channel images, which is more flexible. Using wireless network to transmit data to control mouse movement, using infrared data to control TV station switching and volume control, replacing the function of remote controller (Figure 1).

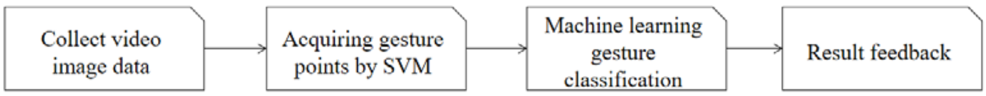

Figure 1 Gesture recognition framework based on SVM

For gesture recognition based on vision, the collected gesture images may have some factors that affect feature extraction, such as uneven illumination, background and noise. These factors will cause the features extracted from gesture images to be unable to describe gestures effectively, or to obtain effective gesture features, thus affecting the accuracy of gesture recognition. The decision function of support vector machine algorithm is determined by a small number of support vectors calculated from training samples, and the complexity of the algorithm depends on the number of support vectors rather than the spatial dimension of samples.

2.3. Feature Extraction

The generated distance histogram may not be very accurate compared with the template of each gesture. To solve this problem, the generated histogram will be adjusted according to the correlation between the calculated histogram and each gesture reference histogram. The main idea is to gradually expand the radius of the surrounding sphere by incremental iteration. In the expansion process, the sample farthest from the current center of the sphere is added into the surrounding sphere. If the binary image or gray image is regarded as a two-dimensional density distribution function, the moment technique can be applied to image analysis.

Gesture energy reflects the intensity of various gestures, and its calculation method is shown in formula (1) [5]:

$$GE = \sum_{i=O}^{P} (|O_{nXi}| + |O_{nYi}| + |O_{nZi} - g|) \tag{1}$$

Among them, O, P is the starting point and end point of gesture; $O_{nXi}, O_{nYi}, O_{nZi}$ is the acceleration of X, Y, Z axis of gesture in user coordinate system.

The acceleration signal duration of gesture action is usually about 1s. In this paper, the mean value is selected as the key feature of each frame acceleration signal. The mean value is the average value of the amplitude of the sample points of a frame acceleration signal:

$$mean = \sum_{i} o(i) / \lambda \tag{2}$$

$o(i)$ is the sampling value of acceleration at the i-th time; λ is the window length. Considering the triaxial acceleration signals comprehensively, one-dimensional features are extracted from each axis to obtain three-dimensional features. For a gesture acceleration data containing 60 frames, 60× 3-dimensional features can be obtained.

The input gesture image can be denoised, and the processing methods are different for different noises in the input image. Common Gaussian noise (Gaussian noise) is usually removed by Gaussian filter. They use different functional formulas, which will directly affect the final support vector of sample training. There is no fixed method for selecting kernel function, and an appropriate kernel function can be selected according to the experimental results through experience.

3. Experimental Data and Analysis

To realize gesture control browser, the basic functions of browser must be completed first [6]. The basic module of browser includes browser initialization, URL input, forward and backward, etc. As a browser plug-in, the gesture module is a supplement to the basic functions, so it is necessary to communicate with the basic module. Based on the bone data provided by Kinect, the trajectory of the joint points of bones of both hands is recorded, processed, and finally recognized by SVM.

The test platform of the experiment is IntelCore2 Duo 3． 00Hz. With 4G memory, the programming environment used by the program is Vistual Studio2008 C++ and OpenCV open source image processing library. Image acquisition uses a camera with 18 million pixels. Among them, SVM train is used for SVM sample training, accepting input in a specific format, and finally generating a model file with suffix ".model", which is the trained model. According to the trained model file and the given input (test sample feature file), SVM predict inputs the category corresponding to the predicted test sample. With direction sensor, acceleration sensor and distance sensor, the hardware configuration of this machine is basically the common hardware configuration of intelligent machine at present.

The design of gesture module is mainly divided into gesture recognition and communication with basic module. Defined gestures and operations on slides are shown in Table 1. When the slide is opened, the right hand stretches out to the right for a period of time, then the gesture recognition mode is turned on, and then W can use the defined gesture to control the slide show operation. In this paper, eight gestures of each person are used for training, and the remaining two gestures are used for testing. For each gesture, some of the training gestures will be used in the calculation of equation (1). The above training set is used to train different SVM classifiers so as to realize classification. This classification method limits that the user's gestures must be trained. In the first stage, 300 normal gesture images are selected for recognition, and the recognition results are shown in Table 2. In the second stage, 300 gesture images of abnormal gestures are mixed with 300 gesture images of normal gestures, and 300 gesture images are randomly selected for recognition. The recognition results are shown in Table 3.

Table 1 Gesture control slide

Function	Start showing	Previous page	Next page	Ballpoint pen	lineation	End the screening
Gesture	Raise your right hand	Right hand up	Right hand down	Right hand to right	Fingertips	Raise your left hand

Table 2 Recognition results of normal gestures

Gesture	Raise your right hand	Right hand up	Right hand down	Right hand to right	Fingertips	Raise your left hand
Recognition accuracy (%)	93.1	94.2	98.5	92.1	89.5	97.4

Table 3 Recognition results of mixed normal and abnormal gestures

Gesture	Raise your right hand	Right hand up	Right hand down	Right hand to right	Fingertips	Raise your left hand
Recognition accuracy (%)	95.6	94.6	94.3	98.5	88.4	90.2

Comparing the recognition results of normal gestures with those of normal and abnormal gestures, it can be seen that the average recognition accuracy of normal gestures is 94.13%, while that of normal and abnormal gestures is 93.6%. This shows that SVM features can express normal gestures and abnormal gestures accurately because of their strong adaptability to the environment. Therefore, in the final machine identification process. The recognition accuracy of abnormal gestures is almost the same as that of normal gestures, and has no significant impact on the overall recognition results.

As a browsing gesture module, the input video images are sampled every 100ms and identified. Since the current input does not necessarily contain valid gestures, the current gesture category is given in the form of probability. When the probability is greater than 95% and one of the given four gestures is determined to be valid for more than five consecutive times, the corresponding browser control command is executed. The system uses kinect as a part of the human-computer interaction model of the natural interactive interface, and carries out experiments through Kinect's Window SDK, and tests the function and performance of the corresponding index parameters of the system. The function Get Tick Count is mainly used in the test, which returns the number of milliseconds elapsed since the operating system started, and its return value is DWORD. At the same time, we can use it for delayed testing. According to the specific conditions of this design, we measured the interaction delay (5 times in total), as shown in Table 4:

Table 4 Test of interaction delay in system implementation

Number of tests	Interaction delay (ms)	Traditional method delay (ms)
1	54	187
2	66	155
3	57	144
4	70	196
5	49	191

It can be seen from Table 1 that the interaction delay in the system operation is in milliseconds. After calculation, the average interaction delay is 59.2msm, which is 3 times higher than the traditional method.

4. Conclusion

Along the way, the discipline of artificial intelligence has been developing and expanding. Although the figure of intelligent robot is still hard to see in our life, with the continuous efforts of many scientists in the future, intelligent robot will develop rapidly. Through the gesture recognition test on Kinect platform, the results show that the gesture recognition algorithm based on SVM vector machine greatly improves the performance of gesture recognition and reduces the interaction delay time compared with the traditional methods. Moreover, the gesture recognition results have better accuracy and accuracy, which greatly improves the gesture recognition rate and reduces errors.

References

[1] Ma, Shuhui., Xia, Bin., Yang, Wenlu., et al. Leap Motion gesture recognition based on SVM. Modern Computer: Mid-Term Journal, no. 23, pp. 55-58, 2017.

[2] Li, Yangqun., Zhou, Mei. Gesture recognition system based on machine learning and its application on mobile terminals. Software Guide, vol. 016, no. 008, pp. 153-155, 2017.

[3] Fang, Lu., Xu, Lirui., Dong, Yanzhi. Research on multi-user tactile gesture recognition based on machine learning. Electroacoustic Technology, vol. 42, no. 07, pp. 28-32, 2018.

[4] Fang, Lu., Xu, Lirui., Dong, Yanzhi. Research on Multi-user Tactile Gesture Recognition Based on Machine Learning. Electroacoustic Technology, vol. 042, no. 007, pp. 28- 32, 2018.

[5] Bao, Zhiqiang., Lu, Chengang. Real-time gesture recognition based on Kinect. Progress in Laser and Optoelectronics, vol. 055, no. 003, pp. 225-230, 2018.

[6] Sun, Peng. Realization of gesture recognition based on pca+svm algorithm. China New Communications, vol. 021, no. 007, pp. 117, 2019.