Available online at http://www.cnki.net

The Journal of China Universities of Posts and Telecommunications

April 2019, 26(2): 30 – 42

http://jcupt.bupt.edu.cn

Research on unified recognition model and algorithm for multi-modal gestures

Guo Xiaopei[1,2], Feng Zhiquan[1,2] (✉), Sun Kaiyun[1,2], Liu Hong[3], Xie Wei[4], Bi Jianping[5]

1. School of Information Science and Engineering, University of Jinan, Jinan 250022, China
2. Shandong Provincial Key Laboratory of Network-based Intelligent Computing, Jinan 250022, China
3. School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China
4. School of Information and Electrical Engineering, Harbin Institute of Technology at Weihai, Weihai 264209, China
5. Affiliated Hospital of Shandong University of Traditional Chinese Medicine, Jinan 250022, China

Abstract

In gesture recognition, static gestures, dynamic gestures and trajectory gestures are collectively known as multi-modal gestures. To solve the existing problem in different recognition methods for different modal gestures, a unified recognition algorithm is proposed. The angle change data of the finger joints and the movement of the centroid of the hand were acquired respectively by data glove and Kinect. Through the preprocessing of the multi-source heterogeneous data, all hand gestures were considered as curves while solving hand shaking, and a uniform hand gesture recognition algorithm was established to calculate the Pearson correlation coefficient between hand gestures for gesture recognition. In this way, complex gesture recognition was transformed into the problem of a simple comparison of curves similarities. The main innovations: 1) Aiming at solving the problem of multi-modal gesture recognition, an unified recognition model and a new algorithm is proposed; 2) The Pearson correlation coefficient for the first time to construct the gesture similarity operator is improved. By testing 50 kinds of gestures, the experimental results showed that the method presented could cope with intricate gesture interaction with the 97.7% recognition rate.

Keywords Kinect, data glove, multi-modal gesture, gesture interaction

1 Introduction

According to the object of recognition, gesture recognition can be divided into static gesture recognition and dynamic gesture recognition. The two types of gestures are fused and identified so that the complex gesture interaction can be done.

Received date: 07-08-2018

Corresponding author: Feng Zhiquan, E-mail: ise_fengzq@ujn.edu.cn

DOI: 10.19682/j.cnki.1005-8885.2019.1004

On the method of recognition, traditional gesture recognition technologies mainly include gesture recognition based on wearable devices and gesture recognition based on red-green-blue (RGB) computer vision of ordinary cameras. Gesture recognition based on wearable equipment refers to the gesture recognition method using data glove or three-dimensional equipment, which limits the natural human-computer interaction to some extent. The traditional vision-based gesture recognition technology is susceptible to interference from complex backgrounds and illumination, and the recognition effect of the algorithm is severely degraded. In recent years, more and more gesture recognition researchers have started to use the combination of depth information and data gloves for more accurate gesture tracking and detection [1]. Since the two are not affected by the illumination changes of the environment and are not linked to the shadow around the object, and they have good robustness in target detection and tracking. In order to improve the recognition rate of gesture recognition based on wearable devices and multi-device fusion, the similarity of the same gesture data format could be used for gesture recognition is proposed. In the preprocessing phase, all the hand gesture data were fitted, that is, smoothing the data to solve the 'jitter' problem. Then, the consistent trend curves were divided into the same gesture according to the shape characteristics of the gesture curves. And the Pearson correlation coefficient was utilized to compare the similarity of the curves. Furthermore, the problem of gesture recognition was transformed into a matching problem of similar curves.

2 Related work

Gesture recognition can be roughly divided into non-visual gesture recognition and visual gesture recognition. Foreign countries have started early in the field of gesture recognition. In terms of non-visual recognition methods, in 1893, Grimes of Bell laboratory first acquires the patent of 'data glove'. Using typical sensor devices such as data gloves, Lee et al. [2] of Carnegie-Mellon University finish a gesture control system of manipulating robots. Ref. [3] reports real-time recognition of Indian and American Sign Language alphabets and numbers based on hand kinematics assessment through an indigenously developed data glove. Kazuma et al. [4] developed a finger motion skill learning support system using data gloves. This system helps a learner to recognize motion errors intuitively by himself/herself by overlaying skilled person's three dimensions (3D) models with learner's 3D models of hands and fingers. And in the field of medical applications, Dutta et al. [5] develop a smart data glove based diagnostic device for better treatment of stroke patients by providing timely estimation of their grasp quality.

In field of recognition methods based on depth information, some researchers take the three-dimensional skeleton data of the human body as the basis and the angle and position of the joint points as the hand gesture feature to recognize hand gestures. For example, Yao et al. [6] propose a gesture recognition method based on RGB-depth (RGB-D) images, which reduces the complexity of gesture matching by using 3D hand contour features. The method of Refs. [7 – 9] is more suitable for small-scale periodic changes or static gesture recognition. On the specific recognition algorithm, Ding et al. [10] propose that they effectively hybridize the 20 main human joint positions captured by the Kinect camera and the joint angle information of 12 critical joints. They use dynamic time warping (DTW) for gesture recognition and get 87.5% recognition rate. Wu et al. [11] describe a novel method called deep dynamic neural networks (DDNN) for multi-modal gesture recognition and a semi-supervised hierarchical dynamic framework based on a hidden Markov model (HMM) is proposed for gesture segmentation and recognition. Yang et al. [12] divide the gesture into eight sub-image areas along the main direction through gesture segmentation, and achieve an overall recognition rate with 95% by Hausdorff-like distance method. Singha et al. [13] propose to recognize isolated gestures by using the difference of gesture speed. A two-stage velocity normalization process using DTW and Euclidean distance-based techniques is introduced, and the decisions of the individual classifiers are combined using classifier fusion model. An accuracy of 94.78 % is achieved using the classifier fusion technique. Lee et al. [14] use Kinect to obtain the gestures of users and recognize gestures through support vector machine (SVM) algorithm [15] to achieve that human can control the smart television (TV) by gestures. Dias et al. [16] utilize HMM [17] to extract and classify gestures and achieve human-computer interaction, but the direct conversion of gesture to operation command reduces the accuracy and naturalness of interaction. What's more, Wang et al. [18] propose the recognition algorithm of dynamic and combined gestures, which based on multi-feature fusion, they establish SVM and HMM model for gestures recognition, the algorithm can overcome the influence of skin object, multi-object moving and hand gestures interference in the background. It can be seen that most of the methods are basically focused on deep learning and template matching, deep learning requires a large amount of sample data, and the template matching method also has the problem of identifying samples that rely too much on the template. Then the gesture recognition method based on curve similarity has come into being. Shehu et al. [19] propose a comparison method of curve representation and similarity measurement algorithm, which reach 97% recognition rate for 30 kinds of dynamic gestures. Shen et al. [20] propose an improved multi-touch gesture recognition method based on the quadratic Bezier curve. The curve parameter equation can be obtained by using quadratic curve, and the property of quadratic Bezier curve is analyzed by utilizing the invariants of the quadratic curve. In the light of the obtained equation, the core three points in the quadratic Bezier curve can be acquired. The algorithm is applied to the vehicle electronic devices based on the Android system. The recognition method based on curve similarity has the advantages of low error and high speed, and depends on simple data structure, which has certain potential in the field of gesture recognition.

Although gesture recognition has been studied extensively at home and abroad, there are some problems to be solved, including real-time, light, tracking under complex motion and occlusion. And, there still exist two critical issues for multi-modal gesture recognition: how to select discriminative features for recognition and how to fuse features from different modalities [21]. Liang et al. [22] also have done some work in this area. They proposed a novel approach to multi-modal gesture recognition by using skeletal joints and motion trail model. In order to avoid the impact of external factors on the gesture recognition and improve the accuracy based on non-visual gesture recognition, gesture trajectories based on the Kinect depth camera and the movement data of hand joints by data glove were obtained. The data were prepossessed and gesture recognition was performed using the feature of gesture curve.

3 Input devices

Various types of gesture recognition for multi-device fusion is focused on the research, so the input devices include Kinect and data glove.

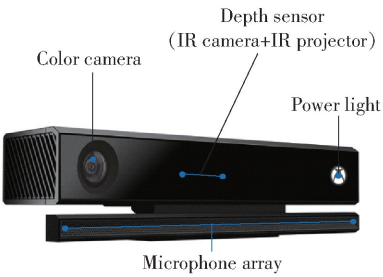
3.1 Kinect sensor

Kinect is the somatosensory peripherals developed by Microsoft for the XBox 360 game consoles. It can simultaneously obtain RGB and depth data, support real-time body and skeleton tracking, and identify a range of actions.

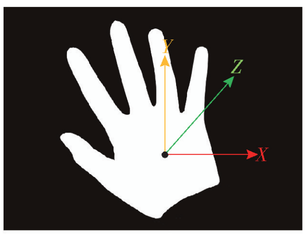
In general, the hand that constantly in motion is present as a trajectory in the parameter space of the gesture model. Depending on the above characteristics and functions of Kinect, the algorithm used Kinect 2.0 (Fig. 1(a)) to lock the 3D coordinates of the center of mass of right hand and track it's position (including depth information), so as to obtain the trajectory data of the center of mass of the right hand, and the three-dimensional coordinate distribution is shown in Fig. 1 (b). This type of trajectory gesture can be detected and identified by dividing the motion trajectory into corresponding subsets.

3.2 Data glove

Data glove is a type of multi-mode virtual reality hardware. The data glove used in this study is shown in Fig. 2(a), it can transform the wearer's action into quantified data or signals through the sensor system composed of multiple sensors. The data were processed to realize gesture recognition. The biovision hierarchy (BVH) data of the movement angle of the hand joints were obtained by data glove, which can be used to extract useful data for the needed finger joints in order to complete the follow-up work. Static and dynamic gestures can be regarded as the movement of 15 joint points obtained by data glove in this study, and the corresponding conditions between its serial number and joints are as follows, 1 – 3: three joints of thumb from bottom to top; 4 – 6: three joints of index finger from bottom to top; 7 – 9: three joints of middle finger from bottom to top; 10 – 12: three joints of ring finger from bottom to top; 13 – 15: three joints of little finger from bottom to top. The actual distribution is illustrated in Fig. 2(b).



(a) Appearance of Kinect 2.0



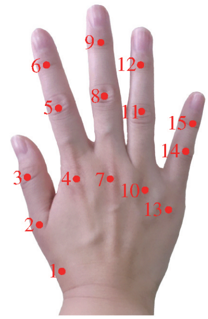
(b) Coordinate distribution of hand centroid

Fig. 1 Kinect and coordinate distribution of hand centroid

4 Recognition algorithm based on multi-modal gesture

Aiming at the phenomenon that there were few kinds of gesture recognition methods for multi-device fusion and the recognition rate was relatively low, a new gesture recognition method was proposed. The following Fig. 3 is the framediagram of this algorithm.

(a) Data glove



(b) Distribution of hand joints

Fig. 2 Diagram of data glove and hand joints

4.1 Curve fitting

The angle change data of the finger joints and the coordinate change of the centroid of the right hand can be regarded as the point sets in the coordinate system. If these data are utilized for subsequent identification, the change of some abnormal data due to handshake will have some impact on the recognition results. Thus a curve fitting method to smooth the data in order to solve the 'jitter' problem is proposed. So the angle change data of each joint were fitted as a curve for the data of finger joints, and the static gesture can be seen as a dynamic gesture with different joints in some stationary states (bent or erect). Similarly, for trajectory data of the centroid of the right hand, the

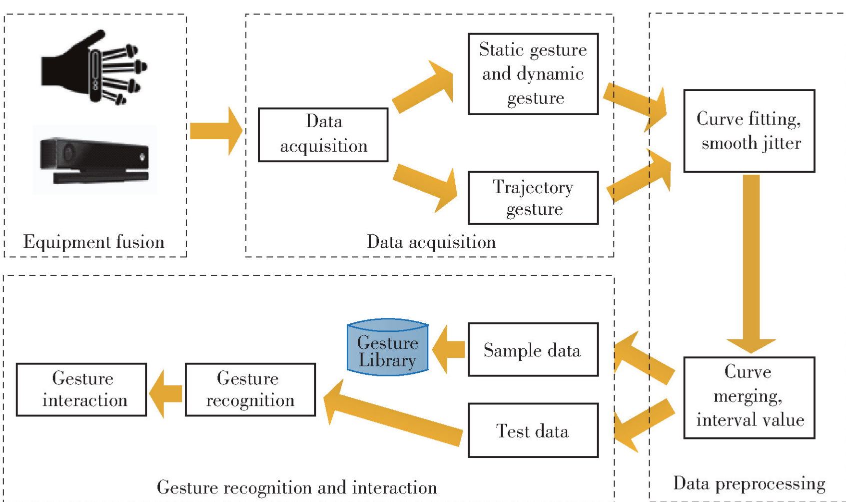


Fig. 3 Gesture recognition frame

change in three directions was also fitted to a curve respectively. Therefore, in order to characterize a complete gesture, multiple curves representing a gesture were connected in a certain order. Each of the composite curves can represent a complete dynamic or static gesture, indicating the trend of the gesture from beginning to end.

The algorithm was based on polynomial curve fitting of least square method, and the basic idea of it is that the fitted curve minimizes the sums of squared error between all data points and estimated points (error is the vertical distance between the data point and the estimate point if it is the two dimensional graphics) [23]. $p(x)$ is called a fitting function or a least square solution in Eq. (1), and

$$p_n(x) = \sum_{k=0}^{n} a_k x^k \tag{1}$$

4.2 Results of curve fitting

A set of data corresponding to one joint (the number of frames captured), which were distributed in a rectangular coordinate system. And it can be found that they presented some ‘trend’. According to this trend, the number of fitting polynomials can be determined and polynomial coefficients can be returned. All the motion data contained in given gestures were fitted to curve, and each serial of joint data was a curve, which indicated the change trend of motion angle from the beginning to the end of the corresponding joint.

The least squares method was used to fitting polynomial curves, the fitting curves were more and more smooth and closer to the real value with the increase of polynomial order, but there would be the over-fitting phenomenon as long as the order exceeded the required maximum value. For instance, Fig. 4 illustrates the fitting curve of the 12th joint in the action of release with three fingers, and their number of times of fitting polynomials are tested 2, 5, 8, 15 respectively, the order here is a random selection in the process of parameter adjustment.

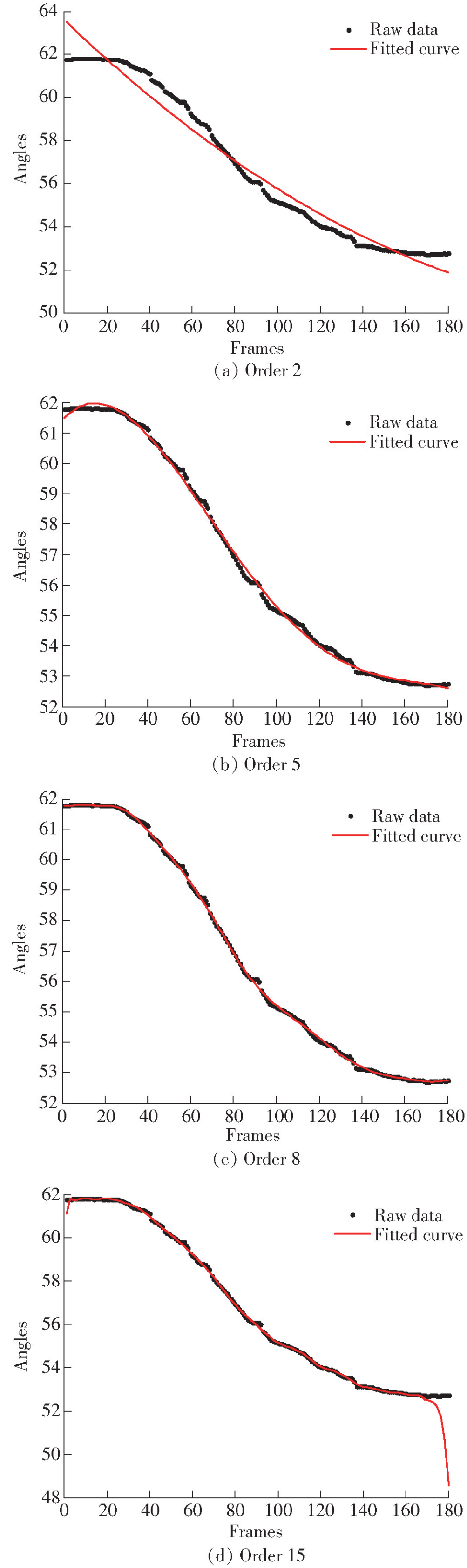
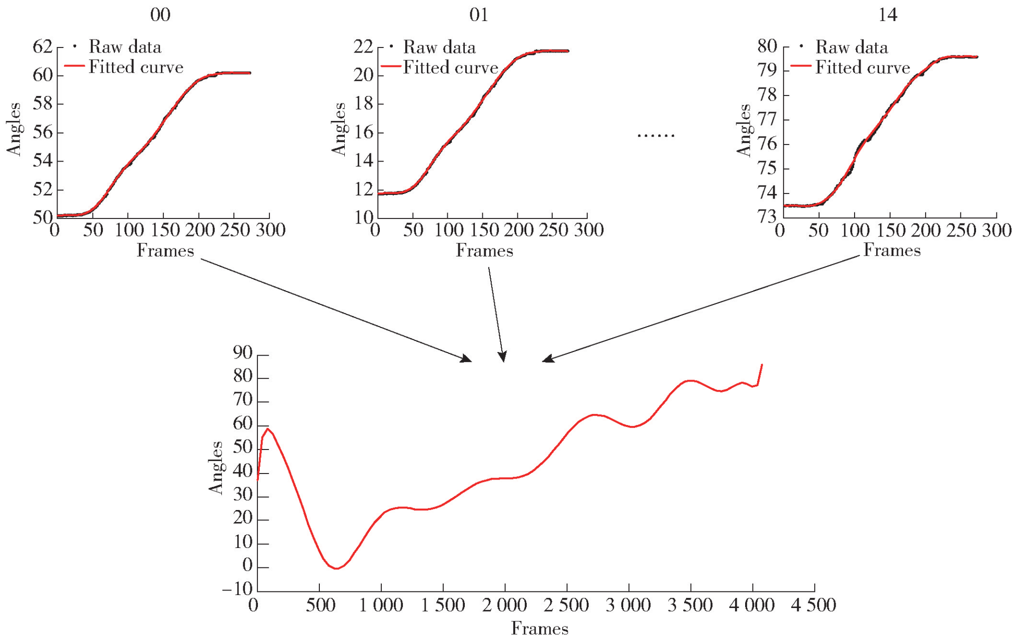


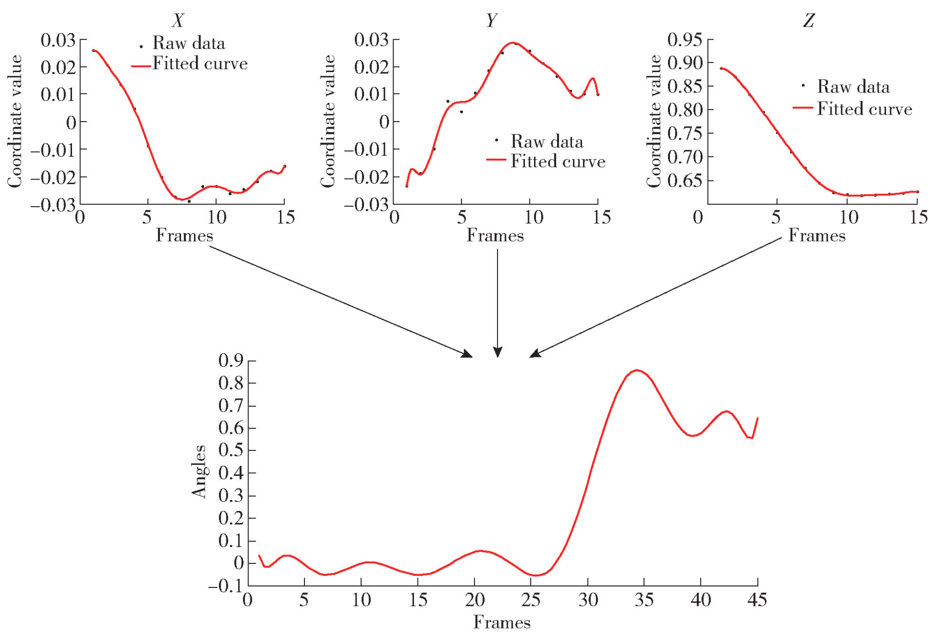
Fig. 4 Different fitting curves for different order numbers

The polynomial order of the experiment is 8, and the fitting effect is best at this time, that is, the sum of square of error between the curve value and the true value is minimum, furthermore there is no over-fitting phenomenon.

As the data obtained were the movement of the 15 joints and the change of the centroid in three-dimensional coordinates. In order to characterize the change of the movement of the whole hand and distinguish trajectory gestures that changing in different directions, each group of data (one joint or one direction for a group) was now connected sequentially. Fig. 5 gives two different gestures to illustrate the process of obtaining a composite curve.



(a) Gesture of grasping with three fingers



(b) Gesture of pushing

Fig. 5 Schematic diagram of gesture curve connection

Therefore, the motion curves of each joint were similar for the same gesture, and the composite curves of all the joint combinations were also approximate, so the preprocessing method can be used to simplify the recognition process and ultimately help gesture recognition.

4.3 Pearson correlation coefficient

The above steps had transformed the problem of gesture recognition into the matching problem of similar curves, then the core of the algorithm was the comparison of the similarity of the curves. In statistics, Pearson product-moment correlation coefficient (PPMCC or PCCs) is used to measure the correlation between two variables. Its value ranges from −1 to 1. The Pearson correlation coefficient method is a statistical method that accurately measures the close degree of two variables.

The Pearson correlation coefficient between two variables is defined as the quotient of the covariance and the standard deviation between the two variables. For the two variables$x$ and y, several sets of data can be obtained through experiments, denoted as (x_i, y_i) $(i=1,2,...,n)$, then the mathematical expression of the correlation coefficient is

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

where $\bar{x}$ and $\bar{y}$ are the mean values of n test values. The range of correlation coefficient r is between −1 and 1, i.e. $|r| \leqslant 1$. The closer to 1 the $|r|$ is, the higher the $x-y$ linear correlation is. If $r=-1$, it is shown that there is a completely negative linear correlation between x and y; and if $r=1$, it is shown that there is a completely positive linear correlation between x and y; if $r=0$, it shows that there is no linear correlation between them [24]. The above equation can also be presented to

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{3}$$

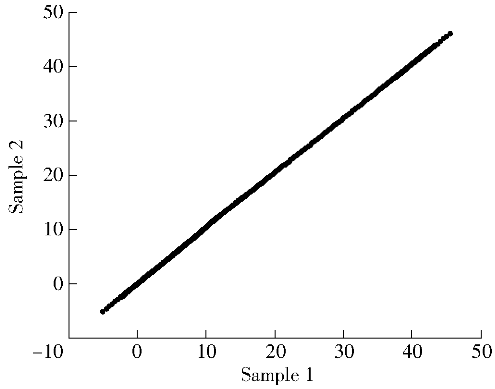
where X and Y are two random variables, $\mathrm{cov}(X,Y)$ is the covariance between the two, and $\sigma_X \sigma_Y$ is the standard deviation between them. After preprocessing the gesture trajectory obtained by Kinect and the movement data of the finger joints obtained from the data glove, the fitting curves of the same gesture were similar, and then two sets of data (sample gesture data and the gesture data to be identified) were treated as the data objects x and y respectively. The Pearson correlation coefficient should be nearly 1 if they are the same gesture. Fig. 6 shows respectively two sets of random curve data, such as static gesture like ‘two fingers open’, dynamic gesture like ‘grasping with three fingers’ and trajectory gesture like ‘drawing rectangle’. The horizontal axis represents the sample data of the gesture, and the vertical axis represents the data to be identified of the gesture.

Due to the difference in the number of frames and the frequency of the acquisition, there is a certain difference in the scalars of the coordinates of each kind of gesture. For the two groups of data, it does not affect the same gesture to present an approximate positive correlation. They show a tendency to tilt to the upper right in a Cartesian coordinate system. This experimentwas used to illustrate the feasibility of adopting correlation coefficients.

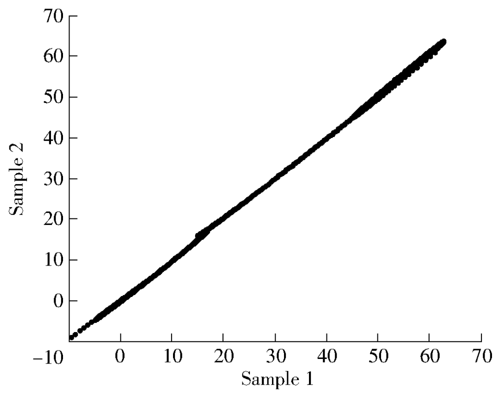
4.4 An improved Pearson correlation coefficient algorithm

Fifty kinds of gestures are defined, including 20 kinds of static gestures, 14 kinds of dynamic gestures and 16 kinds of trajectory gestures. A large number of experiments were carried out with the existing methods and the results as shown in Fig. 7.

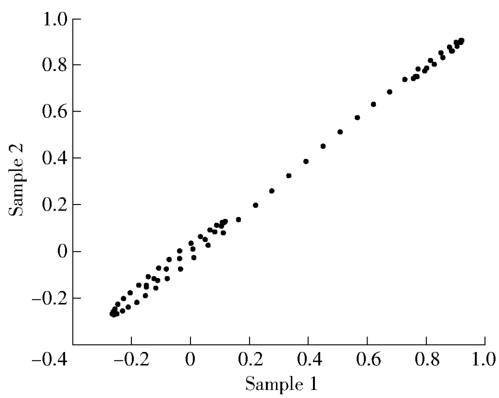
Due to the otherness of different algorithms, there are definite differences in time complexity and recognition rate. It can be observed that most of the current methods have a good recognition of static gestures. But there are some differences between gestures and templates, the recognition of dynamic gestures, especially trajectory dynamic gestures, is not effective. For similar gestures or non-standard gestures, their curves are easy to be confused.



(a) Two fingers open



(b) Grasping with three fingers



(c) Drawing rectangle

Fig. 6 Degree of gesture correlation

Considering the problem of misjudgment in the process of recognition, an improved algorithm was proposed to allocate weight according to the fluctuation intensity of the joint point or a certain direction, that is, the distribution of weight was based on the variance of the motion S_i. Then the similar gesture can be separated, and the algorithm was called 'Pearson correlation coefficient algorithm based on weight', it is referred to as weight-based Pearson correlation coefficient (WPCCs) for short. Eq. (4) is the weight formula

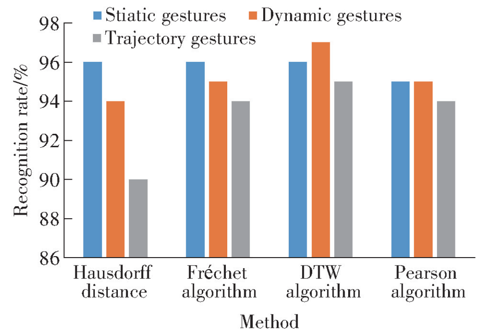


Fig. 7 Comparative experiment of three kinds of gestures under different methods

$$\omega_i = \frac{W_i(\beta)}{\sum_{k=1}^{N} W_i(\beta)} \tag{4}$$

and

$$\beta = \arg\left(\max\left(\frac{S_A(\beta)}{S_B(\beta)}\right)\right) \tag{5}$$

where $W_i(\beta)$ is the function about β, S_A is the variance between classes (between joints or between different directions), which is known from the common variance between classes in images

$$S_A = \omega_0\omega_1(\mu_0 - \mu_1)^2 \tag{6}$$

where ω_0 in Eq. (6) is the ratio of the curve length to the entire curve, ω_1 is the ratio of the remaining curve to the entire curve, μ_0 is the mean value of all values on the segment curve, and μ_1 is the mean value of the remaining curve.

S_B is the within-cluster variance (the same joint or in the same direction). We can see from the general theorem of variance. It is known by the general theorem of variance

$$S_B = \frac{1}{L}\left[\sum_{i=1}^{L}(x_i - \mu_0)^2\right] \tag{7}$$

where L is the number of all values on the segment curve (the method of interval value in this algorithm avoids all data participating in the calculation), x_i is all the values on the segment curve. Then, the improved Pearson correlation coefficient between the two gestures is

$$\rho_{X,Y} = \sum_{i=1}^{N}\frac{\mathrm{cov}(X,Y)}{\sigma_X\sigma_Y}\omega_i \tag{8}$$

Fig. 8 describes the general steps of the WPCCs algorithm in a relatively clear and simple way.

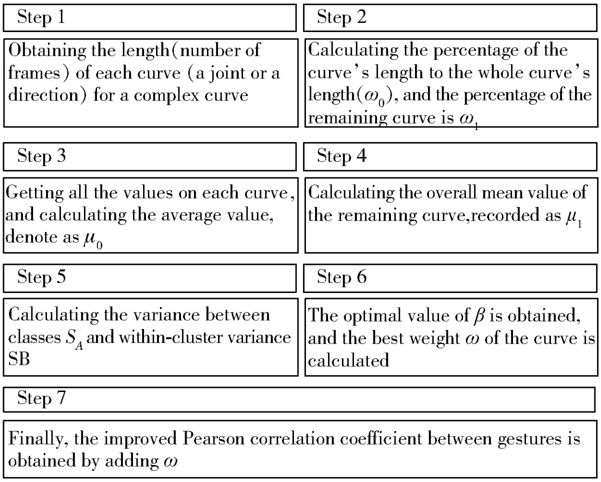


Fig. 8 Gesture recognition algorithm

In a nutshell, this algorithm used the degree of contribution of each segment curve to the entire curve, that is, assigned different weights to it on the basis of the extent of its fluctuation. The method in turn increased the degree of difference between gestures and improved the accuracy rate of the confusing similar gestures or the irregular gestures.

4.5 Algorithm analysis

Based on the original Pearson correlation coefficient and in the case of multi-modal gesture recognition, an improved algorithm WPCCs was proposed, which improved the performance of the algorithm based on a certain degree of time complexity, that is, the accuracy of the recognition of 50 types of gestures. At present, most of the gesture recognition studies mainly focus on gesture detection, gesture segmentation and other works based on image. Such a situation undoubtedly increases the influence of the environment on the gesture recognition results. The object was the real data, which not only avoided the interference of illumination and occlusion, but also improved the recognition speed to a certain extent compared with the method of matching by calculating the distance. While comparing with the existing matching method of correlation coefficients, WPCCs had better recognition effect on the multi-modal gesture. The following experiments evaluate the performance of the algorithm from multiple angles.

5 Experimental results and analysis

The validity of the improved algorithm was validated by the data set of 50 kinds of static gestures, dynamic gestures and trajectory gestures. The recognition result of the algorithm was compared with the traditional algorithm, which showed the superiority of the algorithm.

5.1 Data set of experiments

The data set used in the experiment were collected by the laboratory of 3 subjects, including 20 kinds of static gestures, 14 kinds of dynamic gestures and 16 kinds of trajectory gestures. Each gesture was required to collect 5 times, and a total of 750 samples were included. The sampling frequency and the length of the data were different, and the data set meet the requirement of the contrastive experiment for the different data. Also there were some defects in the traditional algorithm with such data sets. In the statistics of gesture recognition rate, since each gesture had 15 samples, two groups of data were selected randomly as sample data and test data, then the total number of each gesture experiment was 210 times, at this time the recognition rate of the gesture was the number of verification success divided by 210, the result was retained as an integer.

5.2 Result of statistical experiment

The statistical recognition experiments were carried out for 50 kinds of gestures, and the following two tables respectively show the original and the improved correlation coefficients between the gestures in each test set and the sample gestures in the process of recognition (only part of the test data were displayed because of the limitation of space). Other data sources and their preprocessing work were same. Due to different data formats, gesture recognition based on data glove and gesture recognition based on Kinect were performed separately under the control of the threshold range. Therefore, the Pearson correlation coefficient between static gestures, dynamic gestures and trajectory gestures was not compared in the statistical experiments stage.

The analysis shows that in both tables, the Pearson correlation coefficientbetween each gesture and it's same gesture was close to 1, while the similarity to other gestures was randomly distributed between −1 and 1, but there would be a miscalculation, such as the 'Erasure' gesture in Table 1, the Pearson correlation coefficient between the gesture and the same gesture in the sample library was 0.958 2, the similarity between it and the 'Draw a wavy line from left to right' gesture was 0.951 8. Meanwhile, there was a lack of experimental results of other misjudgments because of the space limitation. So the validity and robustness of the original method were weak and the improved algorithm had a certain elevation in these two aspects. From Table 2, the Pearson correlation coefficient between the identical gestures had increased, and it was closer to 1. For different gestures, especially similar gestures or nonstandard gestures, the similarity between them had a reduction to some degree. In short, the algorithm reduced the probability of misjudgment to some extent, enhanced the robustness and improved the recognition rate.

Table 1 Using the original Pearson correlation coefficient algorithm to obtain the degree of correlation between gestures

Gesturename	Opening of the first, fifth fingers	Rotation transformation	Erasure	Release with the second, fifth fingers	Release with two fingers	Draw a wavy line
Opening of the first, fifth fingers	0.996 5	−0.224 0	*	0.635 1	−0.026 2	*
Rotation transformation	−0.218 7	0.940 4	*	0.299 9	0.890 2	*
Erasure	*	*	0.958 2	*	*	0.951 8
Release with the second, fifth fingers	0.662 3	0.187 2	*	0.999 8	0.498 0	*
Release with two fingers	0.025 5	0.854 4	*	0.522 5	0.999 8	*
Draw a wavy line	*	*	0.946 5	*	*	0.998 1

Table 2 Using the improved Pearson correlation coefficient algorithm to obtain the degree of correlation between gestures

Gesturename	Opening of the first, fifth fingers	Rotation transformation	Erasure	Release with the second, fifth fingers	Release with two fingers	Draw a wavy line
Opening of the first, fifth fingers	**0.998 6**	−0.239 0	*	0.592 3	−0.351 2	*
Rotation transformation	−0.247 6	**0.985 3**	*	0.215 5	0.852 3	*
Erasure	*	*	**0.999 8**	*	*	**0.937 8**
Release with the second, fifth fingers	0.634 8	0.103 9	*	**0.999 9**	0.452 1	*
Release with two fingers	−0.012 8	0.833 5	*	0.510 2	**0.999 9**	*
Draw a wavy line	*	*	0.930 2	*	*	**0.998 6**

5.3 Result of comparative experiment

1) Effect of preprocessing on recognition accuracy

The purpose of preprocessing is to smooth the gesture noise data, so with 5 gestures in the gesture database as the identification target, this subsection uses the algorithm to compare the 5 gesture data recognition rate before and after the preprocessing experiment, the experimental results are shown in Table 3.

Table 3 Comparison results of 5 gestures recognition rate under different preprocessing status

Preprocessing status/gesture categories	Open with index finger andmiddle finger/%	Grasp with three fingers/%	Wave with four fingers/%	Draw a rectangle/%	Erasure/%
Before preprocessing	96	95	95	93	94
After preprocessing	**98**	**97**	**98**	**97**	**96**

2) Comparison of random gestures

Ten kinds of gestures were randomly sampled in the gesture data set, which also included static gestures, dynamic gestures and trajectory gestures. Comparing this algorithm with the methods used by others in their works for the recognition rate of these 10 gestures, Table 4 is the recognition result.

Table 4 Comparison results of 10 random gestures recognition rate

Gesture type/method	Hausdorff/%	Fréchet/%	DTW/%	Pearson correlation coefficient/%	WPCCs/%
Fist	94	95	96	96	**97**
Opening with five fingers	95	95	96	95	**97**
Grasp with three fingers	94	95	96	94	**97**
Wave with four finers	95	95	95	95	**96**
Release with the second, third fingers	95	95	96	95	**97**
Underline from left to right	93	94	94	93	**97**
Underline from top to bottom	92	93	94	94	**96**
Release with the third, fourth, fifith fingers	95	95	96	96	**98**
Grasp with two fingers	95	96	96	96	**97**
Draw a triangle	93	93	95	95	**97**
Average	94	95	95	95	**97**

3) Comparison of similar gestures

There are some gestures inthe gesture data set. The change curves were very similar. We call them as similar gesture. Comparing this algorithm with the methods used by others in their works for the recognition rate of these similar gestures, Table 5 is the recognition result.

Table 5 Comparison results of 10 similar gestures recognition rate

Gesture type/method	Hausdorff/%	Fréchet/%	DTW/%	Pearson correlation coefficient/%	WPCCs/%
Grasp with three fingers	93	93	95	94	**98**
Screw with three fingers	94	93	95	94	**98**
Grasp with five fingers	94	94	95	95	**98**
Rotate with five fingers	94	95	96	95	**98**
Draw a wavy line	93	93	94	94	**96**
Draw an arc	94	94	94	94	**97**
Draw a triangle	93	93	95	94	**97**
Draw a circle	93	94	95	94	**98**
Release with the first, second fingers	93	93	95	94	**98**
Release with the first, third fingers	93	94	96	95	**98**
Average	93	94	95	94	**98**

4) Comparison of nonstandard gestures

At present, most of the algorithms have a higher dependence on templates, once one gesture differs from the same gesture in the template or the similarity between gestures and standard gestures is not high, there will be a certain probability of miscalculation. Comparing this algorithm with the methods used by others in their works for the recognition rate of the following nonstandard gestures, Table 6 is the recognition result.

Table 6 Comparison results of 10 nonstandard gestures recognition rate

Gesture type/method	Hausdorff/%	Fréchet/%	DTW/%	Pearson correlation coefficient/%	WPCCs/%
Opening with the second, third fingers	91	92	94	93	**96**
Grasp with five fingers	92	92	94	94	**96**
Opening with the first three fingers	92	94	93	94	**97**
Screw with three fingers	90	92	93	92	**97**
Release with the second, third, fourth, fifth fingers	93	94	94	94	**97**
Draw a rectangle	89	91	91	91	**96**
Draw a wavy line	91	91	92	92	**97**
Nine	94	93	94	94	**97**
Erasure	92	94	93	94	**97**
Withdraw	92	94	95	93	**98**
Average	92	93	93	93	**97**

5) Comprehensive contrast experiment

In order to demonstrate the validity of the proposed method, the algorithm was compared withthe methods used by others in their works for 50 kinds of gestures. In the experiment, all the methods used the unified gesture library. The recognition results are presented in Fig. 9.

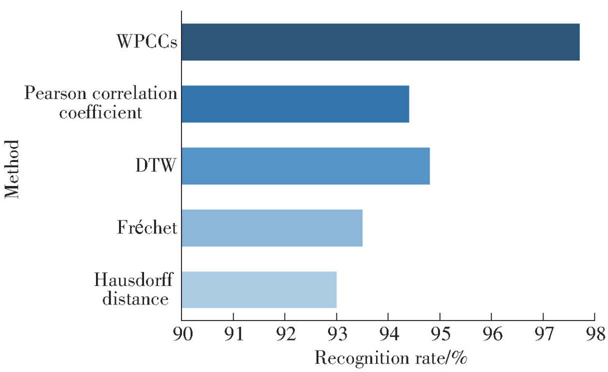


Fig. 9 Comparison of the comprehensive recognition rate of 50 kinds of gestures

5.4 Experimental analysis

The above experimental results firstly showed the validity of the improved Pearson similarity algorithm on the real data set. The algorithm can also act on other data sets with the same type to make its classification decision. Secondly, compared with the traditional Pearson correlation coefficient algorithm, the improved algorithm improved the recognition rate 2% ~ 3% in the randomly selected 10 kinds of gesture recognition, increased the recognition rate by 3% ~5% in 10 kinds of similar gesture recognition, and improved the recognition rate 4% ~ 5% in the 10 kinds of nonstandard gesture recognition. That is, it can distinguish easily confusing similar gestures and irregular gestures, while the improved algorithm had a significant improvement in robustness and applicability compared with traditional distance method.

6 Conclusions

A method of gesture recognition based on multi-device fusion for multi-modal gestures is presented. It including recognition of static gestures, dynamic gestures and based on data glove, and recognition of trajectory gestures based on Kinect. The motion data of the finger joints and the trajectory data of the center of mass of the hand were obtained by data glove and Kinect respectively. The same preprocessing method was used to fitting and disperse all data, and the similarities between the curves were calculated, i. e., Pearson correlation coefficient between the gestures. The correlation coefficient must be approximately 1 for the same gesture. For the most confusing opposite gestures, such as ‘grasp with five fingers’ and 'release with five fingers', 'wave to the left' and 'wave to the right', the correlation coefficient of them was negative or close to −1 due to the roughly symmetrical trend of the curve. For the multi-source heterogeneous data from data glove and Kinect, the same preprocessing method and recognition algorithm were designed, which were processed and identified in real time and in parallel. 97.7% recognition rate and a certain degree of robustness were achieved. For the future work, we will combine more other modals (speech, posture, iris recognition, etc.) to improve the performance of human computer interaction.

Acknowledgements

This work was supported by the National Key R&D Program of China (2018YFB1004901) and the National Natural Science Foundation of China (61472163, 61472232).

References

1. Liu X. Research of dynamic gesture recognition based on the kinect depth image. Shenyang, LN, China: Northeastern University, 2014: 77p (in Chinese)
2. Lee C, Xu Y. Online, interactive learning of gestures for human/robot interfaces. Proceedings of IEEE International Conference on Robotics and Automation, Apr 22 − 28, 1996, Minneapolis, MN, USA: IEEE, 2002(4): 2982 − 2987
3. Nayan M K, Manalee D S. Recognition of sign language alphabets and numbers based on hand kinematics using a data glove. Procedia Computer Science, 2018(133): 55 − 62
4. Kazuma I, Masato S, Hirokazu T. Development of finger motion skill learning support system based on data gloves. Procedia Computer Science, 2014, 35(6): 1307 − 1314
5. Dutta D, Modak S, Kumar A, et al. Bayesian network aided grasp and grip efficiency estimation using a smart data glove for post-stroke diagnosis. Biocybernetics and Biomedical Engineering, 2017, 37(1): 44 − 58
6. Yao Y, Zhang L J, Qiao W B. Hand part labeling and gesture recognition from RGB-D data. Journal of Computer-Aided Design and Computer Graphics, 2013, 25(12): 1810 − 1816 (in Chinese)
7. Li H B, Ding L J, et al. Static three-dimensional gesture recognition based on kinect skeleton data. Computer Applications and Software, 2015, 32(9): 161 − 165 (in Chinese)
8. Batabyal T, Chattopadhyay T, Mukherjee D P. Action recognition using joint coordinates of 3D skeleton data. IEEE International Conference on Image Processing, Sep 27 − 30, 2015, Quebec City, Canada: IEEE, 2015: 4107 − 4111
9. Zhu H M, Pun C M. Human action recognition with skeletal information from depth camera. IEEE International Conference on Information and Automation, July 26 − 31, 2014, Hulun Buir, NM, China: IEEE, 2014: 1082 − 1085
10. Ding I J, Chang C W. Feature design scheme for kinect-based DTW human gesture recognition. Multimedia Tools and Applications, 2016, 75(16): 9669 − 9684
11. Wu D, Pigou L, Kindermans P J, et al. Deep dynamic neural networks for multimodal gesture segmentation and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(8): 1583 − 1597
12. Yang X W, Feng Z Q, Huang Z Z, et al. A gesture recognition algorithm using hausdorff-like distance template matching based on the main direction of gesture. Applied Mechanics and Materials, 2015, 713 − 715(3): 2156 − 2159
13. Singha J, Laskar R H. Hand gesture recognition using two-level speed normalization, feature selection and classifier fusion. Multimedia Systems, 2017, 23(4): 499 − 514
14. Lee W P, Kaoli C, Huang J Y. A smart TV system with body-gesture control, tag-based rating and context-aware recommendation. Knowledge-Based Systems, 2014, 56(3): 167 − 178
15. Ghaleb F F M, Youness E A, Elmezain M, et al. Vision-based hand gesture spotting and recognition using CRF and SVM. Journal of Software Engineering and Applications, 2015, 8(7): 313 − 323
16. Dias T, Variz M, Jorge P, et al. Gesture interaction system for social web applications on smart TVs. Proceedings of the 10th Conference on Open Research Areas in Information Retrieval, May 15 − 17, 2013, Lisbon, Portugal, 2013: 225 − 226
17. Kim Y C, Jiang J S. Remote gesture control using HMM classifier based on electric potential sensors. International Journal of Multimedia and Ubiquitous Engineering, 2015, 10(2): 413 − 420
18. Wang L, Liu G X, Duan H Y. Dynamic and combined gestures recognition based on multi-feature fusion in a complex environment. The Journal of China Universities of Posts and Telecommunications, 2015, 22(2): 81 − 88
19. Shehu V, Dika A. Curve similarity measurement algorithms for automatic gesture detection systems. Proceedings of IEEE 35th International Convention MIPRO, May 21 − 25, 2012, Opatija, Croatia: IEEE, 2012: 973 − 976
20. Shen L, Huo S C, Chen M S, et al. Multi-touch gesture recognition algorithm of vehicle electronic devices-based on Bezier curve optimization strategy. 2017 32nd Youth Academic Annual Conference of Chinese Association of Automation (YAC), May 19 − 21, 2017, Hefei, AH, China: IEEE, 2017: 720 − 723
21. Wu J X, Cheng J. Bayesian co-boosting for multi-modal gesture recognition. The Journal of Machine Learning Research, 2017, 15(1): 3013 − 3036
22. Liang B, Zheng L. Multi-modal gesture recognition using skeletal joints and motion trail model. Workshop at the European Conference on Computer Vision, Sep 6 − 12, 2014, Zurich, Switzerland, 2014: 623 − 628
23. Jia X Y, Xu C S, Bai X. The invention and way of thinking on least squares. Journal of Northwest University (Natural Science Edition), 2006, 36(3): 507 − 511 (in Chinese)
24. Yang F, Feng X, Ruan L, et al. Correlation study of water tree and VLF tanδ based on pearson correlation coefficient. High Voltage Apparatus, 2014, 50(6): 21 − 31 (in Chinese)

(Editor: Ai Lisha)