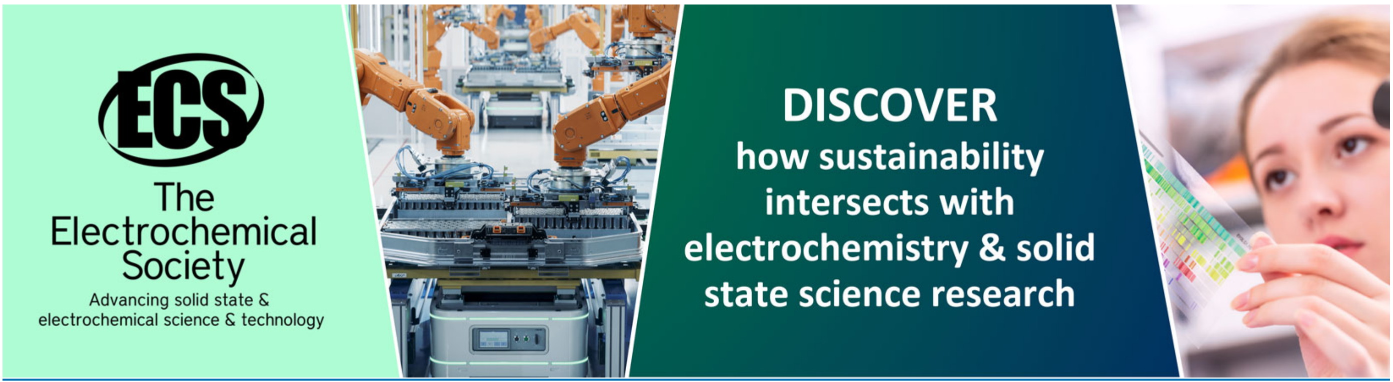
Journal of Physics: Conference
Series

Dynamic Gesture Recognition based on LeapMotion and HMM-CART Model


ECS
The Electrochemical Society
Advancing solid state & electrochemical science & technology
DISCOVER
how sustainability intersects with electrochemistry & solid state science research

Dynamic Gesture Recognition based on LeapMotion and HMM-CART Model

Qixiang Zhang and Fang Deng

College of Computer Science and Technology, Beijing University of Posts and
Telecommunications, Beijing, 100876, China
Email: manatea@bupt.edu.cn, dengfang@bupt.edu.cn

Abstract. This paper focuses on improving the recognition accuracy of dynamic gesture learning, and proposes dynamic gesture track recognition strategy based on HMM-CART model structure. It integrates the modeling ability of the HMM (Hidden-Markov-Model) for time series data and the interpretability of CART (Classification-And-Regression-Tree) for its ability of fast classification and regression, and use it for dynamic gesture recognition. Time series data of movement gestures collected by LeapMotion sensor will be first divided into four channels: finger shape, palm normal vector, palm ball radius and palm displacement vector, and then HMM in HMM Layer will be built for each channel, finally the likelihood probability of model calculated for each sub-sequence of observation should be classified as the input of the CART model in CART Layer to identify the gesture. Experiments show that the accuracy of dynamic gesture recognition method using HMM-CART model is higher than that of traditional single channel HMM.

1. Introduction

In recent years, gesture recognition in many different areas attracted the interest of a growing number of studies, such as human-computer interaction, robotics, computer games, and automatic sign language interpretation and so on. In the development of human-computer interaction which has gone through the interaction from the initial two-dimensional image by simple operator interface to interact more intuitively and naturally, and gesture recognition has also been improved from a wearable device detection to more convenient camera-based visual gesture recognition. Currently more sophisticated visual identification products with high efficiency depth feature extraction technique is widely used in the world: Microsoft's Kinect using both sides of the infrared sensor and the middle of the RGB camera through the optical coding technology to obtain depth information and human skeletal information[1], Intel's Real Sense is similar to Kinect, and its black box model of the package allows users to store features in accordance with established models[2], LeapMotion is innovative in terms of gesture recognition software, unlike Kinect, LeapMotion device is clear for gesture recognition and accurate calculation of finger and hand structure. Compared to the depth of the camera like Kinect and other similar devices, it produces more limited gesture local information and interactive area, the advantage lays in the local data recognition more accurate [3].

Gesture recognition is the process of eigenvalue extraction and gesture model parameter estimation of the image after segmentation of the gesture area. It is a process of mapping a point or track in a parameter space to a subset of the space. It includes static gesture recognition and dynamic gesture recognition, and dynamic gesture recognition can eventually be converted to static gesture recognition. Lathe et al. [4] achieved gesture recognition by calculating the similarity of the hausdorff distance calculation between the target data and the gesture templates obtained from sample gesture training, which is simple and fast, but the degree of division between different gestures depends on the choice of eigenvalues, so the recognizable gestures are more limited. Tusor at al. [5] used a scheme based on artificial neural networks and fuzzy theory to classify different gestures by setting a priori fuzzy rule. Its recognition accuracy based on a large number of training data, although the use of fuzzy network to improve the training iteration speed, but the training process is still too long, which does not apply to real-time recognition. X Wang et al.[6] Used Baum-Welch algorithm to train the HMM (Hidden-Markov-Model), which can handle time-based information of different lengths and has high recognition efficiency for dynamic gestures. However, with the development of interactive applications and the improvement of recognition accuracy requirements, a single HMM has been unable to meet the needs of the application, therefore, XY Wang et al.[7] Proposed a HMM-FNN-based model structure that combines the HMM with a fuzzy neural network to establish a more complex gesture by setting up fuzzy rules. Indeed, it has improved the accuracy, but still has the shortcomings such as the dependence of fuzzy rules on prior experience and insufficient convergence rate due to more hidden layer nodes.

This paper focuses on improving the recognition accuracy of dynamic gesture learning, and proposes dynamic gesture track recognition strategy based on HMM-CART model structure. It integrates the modeling ability of HMM for time series data and the interpretability of CART (Classification-And-Regression-Tree) for its ability of fast classification and regression, and uses it for dynamic gesture recognition. Time series data of movement gestures collected by LeapMotion sensor will be first divided into four channels: finger shape, palm normal vector, palm ball radius and palm displacement vector, and then HMM in HMM Layer will be built for each channel, finally the likelihood probability of model calculated for each sub-sequence of observation should be classified as the input of the CART model in CART Layer to identify the gesture. Experiments show that the accuracy of dynamic gesture recognition method using HMM-CART model is higher than that of traditional single channel HMM.

The rest of the paper is organized as follows: Section 2 introduces the collection and pre-processing of gesture data. Section 3 describes the structure of dynamic gesture model based on HMM-CART, including the structure and training principle of related learning model. Section 4 describes the experiment that introducing the classification and training process of new model and compare it with traditional model according to the experimental results. Finally, Section 5 ends the paper with a summary.

2. Data Acquisition and Preprocessing

Compared with kinect, LeapMotion uses a similar infrared camera technology, and its exquisite design is specifically used to capture the trace of the hand. It can be through the infrared light reflection of the LED tracking about 1 meter in diameter within the coordinates of the coordinates of the hand. Figure. 1 shows that the LeapMotion device captures 215 frames per second and the detection accuracy can reach 0.01 mm. API support cross-platform, allowing access to the original data such as coordinates of five fingers, the strength of grip hand, the hand rotation angle, which help us to build a data set to achieve and analyze the algorithm of gesture recognition. We will have theses tracking data collected and extracted through the information acquirement of the LeapMotion device.

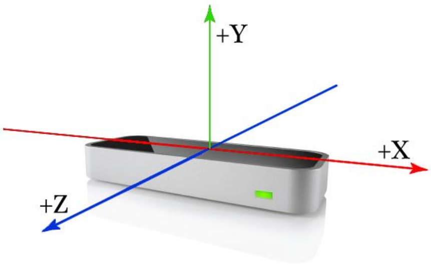


Figure 1 LeapMotion device

The acquisition of gesture data needs to define the start and end of gesture position. Therefore, this paper summarizes a threshold judgment scheme based on hand shape template and gesture movement velocity and direction change rate. We determine starting and ending points of the gesture through whether the gesture movement speed exceeds the threshold, and set the direction change rate to prevent the gesture of the inflection point when the speed may be lower than the threshold. Like gesture V, in acquisition of gesture V, the velocity in bottom inflection point of V may be lower than the threshold.

The definition and extraction of features are related to the accuracy of the overall model training. The LeapMotion device contains a wide variety of features in its original dataset that it allows access to. In order to avoid the high-dimensional observation sequence learning, it is necessary to do dimensionality reduction first. According to the characteristics of hand shape, the dynamic change of the gesture in space could be divided into several channels in three-dimensional space such as trace of palm position, the trace of normal vector of the palm, the change of ball radius holding in palm and the stretch of five fingers. Wherein the variation of each gesture information is represented by a time series of successive characteristics, the Gaussian HMM training by continuous observation of sequence can retain more original information than that by Discrete HMM, and

The Gaussian HMM modeling is carried out for each one-dimensional feature of the above division.

Each gesture detection process may have a range of size, shape distortion and other issues, and each gesture to achieve the shape and size of the space is different. Even the same set of gestures training data its eigenvalues will vary greatly if not get data preprocessed, so we need to normalize the gesture data into the same format under the data. This paper realizes the same specification data mapping by normalizing the eigenvalues of coordinates.

The experiment will be normalized using the L^2 norm, and the norm $x(x_1, x_2 ..., x_n)$ for the vector L^2 is defined as follow:

$$norm\ (x) = \sqrt{x_1^2 + x_2^2 + ... + x_n^2} \tag{1}$$

To normalize x to the unit L^2 norm, as create a mapping from x to x' so that the norm x' of L^2 is 1, as shown in (2):

$$1 = norm(x'\quad) = \frac{\sqrt{x_1^2 + x_2^2 + ... + x_n^2}}{norm(x)} = \sqrt{x'^2_1 + x'^2_2 + ... + x'^2_n} \tag{2}$$

Which is:

$$x'_i = \frac{x_i}{norm\ (x)}$$

The normalized gestational feature observation sequence can be expressed as: $m = \{x_i \mid i \in [1, n], x_i \in [-1,1]\}$

3. Model Design and Training

The dynamic gesture modeling method of HMM-CART model proposed in this section preserves the ability of time model modeling based on HMM, and also makes full use of the rule interpretability of CART model and fast classification and regression ability.

3.1 HMM-CART Model

As shown in Figure. 2, the HMM-CART model is divided into four layers: Input Layer, HMM Layer, Mid Layer, and CART Layer. In Input Layer, the data is reduced dimension and normalized by the feature, and then divided into four channels: finger shape, palm normal vector, palm ball radius and palm displacement vector, where each channel contains Part of the information of a gesture that will be modeled as a separate observation sequence into HMM Layer according to the design of the different gestures.

The HMM is similar to the traditional HMM, but the difference is that the HMM in the HMM Layer only covers a certain kind of information of the gesture, and the traditional HMM does not carry out the multi-channel division of the feature.

Each channel in HMM Layer is concerned only with which models are associated with the sequence, for example, the gesture sequence feature is based on the four observed sequences of s_v , s_r , s_n , s_e , Where s_v is the displacement vector sequence, s_r is the palm ball radius sequence, s_n is the palm normal vector sequence and s_e is the finger extension degree sequence, each data sequence has a corresponding HMM model. We can evaluate the observed sequence values by entering the HMM and then enter the score results into the next layer. At HMM Layer, each node represents the HMM of a gesture under that channel. The total number of HMM nodes is $n_g \cdot n_c$, where n_g represents the number of design gestures and n_c represents the number of channels.

The middle layer evaluates the likelihood probability of each model as a set of likelihood probability features and classifies the eigenvector into CART Layer after performing the normalization algorithm.

As the input root node of CART Layer, the decision tree as a binary tree structure contains several decision points and the result nodes, which can be easily, expressed as if-then rules. Starting from the root decision point, enter the data for two sub-program decision-making calculation, and ultimately reach the output of a result node. The output layer contains all possible classification results and integrates the results as the final output of the model.

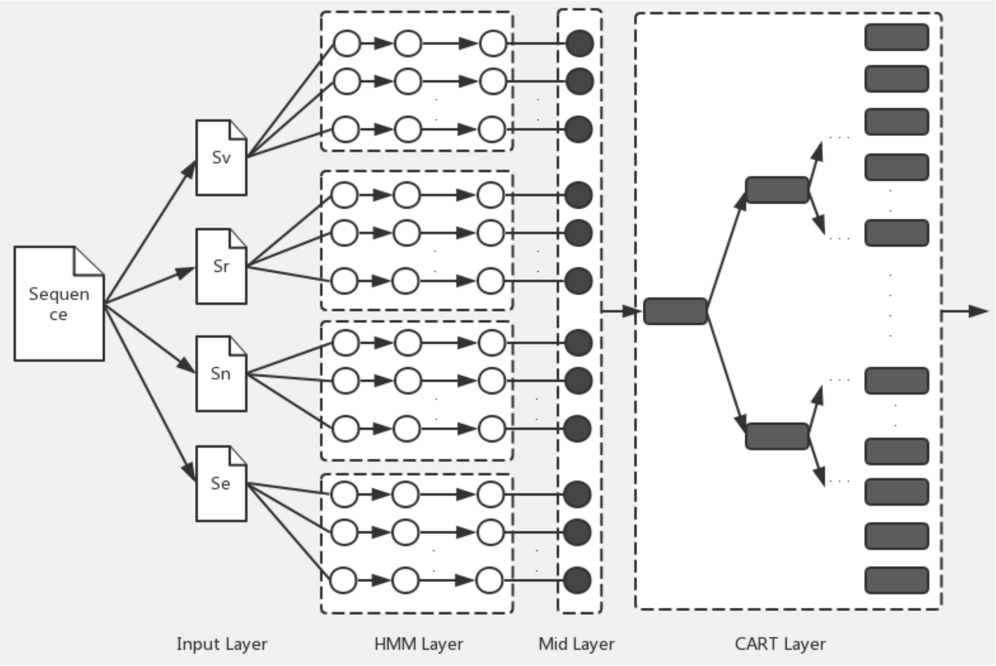


Figure 2 HMM-CART model

3.2 Model Training

Hidden-Markov-Model is distinguished in timing-based pattern recognition applications such as gestures, voice, handwriting, and semantics. The HMM can map the transition probability between the observed and implicit states whose implicit state is invisible and the output of the observed state that is dependent on the implicit state is visible. There is a transition probability between each implied state, and the transition probability of the implied state to the observed state is obeying the distribution, thus a HMM can be represented by the implicit state transition probability matrix and the confusion matrix as well as the initial state probability matrix.

Training an HMM model is the process of giving an observation state, finding the optimal state transition matrix and confusing the matrix. In the training process, by continuously updating the HMM parameters the $P(O \mid \lambda)$ value reaches a local optimum. The training algorithm uses the Baum-Welch algorithm, which will iterate over the model parameters according to the maximum likelihood estimate.

We assume that the initial HMM parameter is $\lambda = (A, B, \pi)$, first calculate the forward variable α and the backward variable β as shown in follow formula:

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, \lambda) \tag{3}$$

$$\beta_t(j) = P(O_{t+1} O_{t+2} \dots O_T \mid \lambda) \tag{4}$$

Where the forward probability α is the sum of all the probabilities reaching the state i at time t, and the backward probability β is the probability of partially observing the sequence $O_{t+1} O_{t+2} \dots O_T$ when the state i is reached, so $P(O \mid \lambda)$ can be further expressed as:

$$P(O \mid \lambda) = \sum_{i=1}^{N} \alpha_t(i) \beta_t(i), 1 \le t \le T \tag{5}$$

According to (5), we define the expectation ξ and γ , where ξ is the probability that the state q_t is i at time t and the state q_{t+1} is j when time is t + 1:

$$\xi_t(i, j) = \frac{P(q_t = i, q_{t+1} = j \mid O, \lambda)}{P(O \mid \lambda)} \tag{6}$$

We define γ as the probability of the state i at t when given the observed state sequence and the HMM parameter:

$$\gamma_t(i, j) = P(q_t = i \mid O, \lambda) \tag{7}$$

Finally, we use (6) and (7) to re-iterate to calculate $\lambda = (A, B, \pi)$, the formula is as follows:

$$\overline{\pi_i} = \gamma_1(i), 1 \le i \le N \tag{8}$$

$$\overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \xi_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)}, 1 \le i \le N, 1 \le j \le N \tag{9}$$

$$\overline{b_{ik}} = \frac{\sum_{t=1, \alpha_t = k}^{T} \gamma_t(i)}{\sum_{t=1}^{T} \gamma(j)}, 1 \le j \le N, 1 \le k \le M \tag{10}$$

It is possible to obtain a maximum likelihood estimate of the HMM model after a number of iterations. The maximum likelihood estimation of the HMM model can be obtained by iteratively calculating the above three equations.

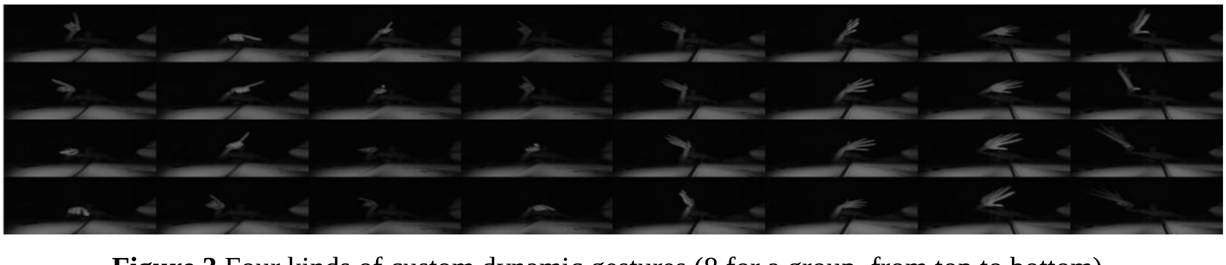


Figure 3 Four kinds of custom dynamic gestures (8 for a group, from top to bottom)

The best segmentation of the classification regression tree is based on the feature realization of the model dataset. It constructs the prediction criterion by constructing the binary tree, which can deal with large amount of data and high data effectively, and can also deal with the continuous variable data in this paper. CART is based on the Gini impurity to make the segmentation decision. For continuous target variable, the minimum variance is used as the splitting rule to generate the binary tree. For a set of data $D = \{(x_1, y_1), (x_2, y_2),..., (x_n, y_n)\}$, we define two sets of regions to be segmented:

$$\begin{cases} R_1(j,s) = \{x \mid x_j \le s\} \\ R_2(j,s) = \{x \mid x_j > s\} \end{cases} \tag{11}$$

The selection rule for the partitioned region must satisfy the sum of the weights of the two regions. So that each partition of area data have to ensure relative polymerization to achieve the purpose of training.

Over fitting may be presented during the binary state of the training process, since the characteristics of the model training samples contains too precise and noise information, it is necessary to prune the tree generated by the method of cross-validated with the validation data set, and by adding a penalty factor to select the high error rate for branch reduction to solve the problem of fitting.

4. Experiments

In this paper, the CPU for the Intel Core i5-3470 3.20GHz, 8G memory for the ordinary PC deployment of the gesture recognition experiment, as shown in Figure. 3 gesture definition is divided into four kinds of fingers: draw round, draw fork, left slide, and right slide. Each gesture collected 100 data, which divided into four channels as follows: displacement vector sequence, palm ball radius sequence, palm normal vector sequence and finger extension degree sequence. Total 1600 data that division of the four different channels for a group of 400 groups, will be in accordance with the ratio of 8: 2 cross validation.

In addition to the definition of simple gestures in the experiment, such as round, fork, but also defines a more complex left slide, right slide gestures. As is shown in Figure. 3, the first two kinds of gestures are the process of the extension of index finger. The latter two gestures are the level of the process of placing from the left or from the right in the opposite direction with five fingers stretch.

First, during the training the sequence of displacement vector, palm ball radius, palm normal vector and finger extension degree will be extracted for HMM training. And then the HMM-CART classification regression of the likelihood of each HMM will have completed.

The training process of the experiment is described in the previous section, and the training is divided into two phases: Firstly, 80 × 4 sets of data are input into the HMM model for parameter training. Then, 320 sets of likelihood probability vectors are obtained by inputting the training data of each training set into the HMM as training input train of CART model.

Figure. 4 shows a change in the palm displacement characteristics of a left slide and a right slide gesture. It can be seen in Figure 4 that although the Y and Z axes have a slight change, the X axis completely covers the gesture information.

Finally, the HMM-CART model and the traditional HMM model were predicted by 80 sets of test data. The test results are shown in Table 1.

It can be seen from the experiment that the multi-channel decomposition of the gesture feature can make the recognition effect of the HMM for the single channel feature significantly improve, and still maintain a high recognition rate after classification and aggregation.

Table 1 Comparison of HMM - CART Model and Traditional HMM Model Recognition Rate

Hand Gesture Recognition Model	Circle	Cross	Left Slic	Right Slide
Recognition rate of Conventional HMM (%)	91.3	76.6	92.6	80.8
Recognition rate of HMM-CART (%)	95.7	93.3	98.1	95.6

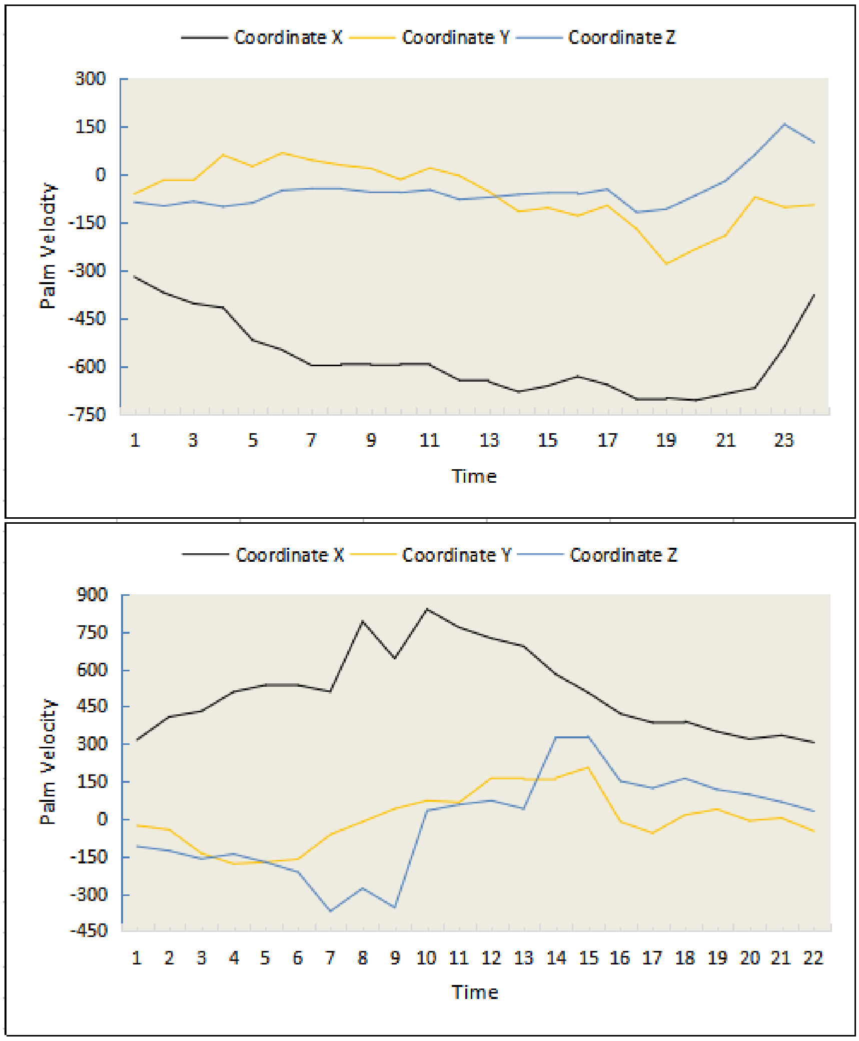


Figure 4 Left slide (top) right slide (bottom) palm displacement characteristics

5. Conlusion

In this paper, HMM-CART model is proposed for modeling and recognizing dynamic gestures. First, it uses the threshold acquisition method to collect the gesture information through the LeapMotion sensor, and then split it into four component channels and normalize the timing data as the input data of the HMM-CART model.

It integrates the modeling ability of the Hidden-Markov-Model for time series data and the interpretability of Classification-And-Regression-Tree for its ability of fast classification and regression, and uses it for dynamic gesture recognition.

In the future, we will collect data through multiple LeapMotion devices and integrate gesture implicit features to improve the recognition of gestures.

6. References

[1] Zhang Z. Microsoft Kinect Sensor and Its Effect[M]. IEEE Computer Society Press, 2012.

[2] Lin C P, Wang C Y, Chen H R, et al. RealSense: directional interaction for proximate mobile sharing using built-in orientation sensors[C]// ACM International Conference on Multimedia. ACM, 2013:777-780.

[3] Marin G, Dominio F, Zanuttigh P. Hand gesture recognition with leap motion and kinect devices[C]// IEEE International Conference on Image Processing. IEEE, 2015:1565-1569.

[4] Latha P S, Babu M R, Siva K, et al. Hand Gesture Recognition using skeleton of Hand and Hausdroff Distance[J]. International Journal of Advanced Research in Computer Science, 2011,

2(2):346-354.

[5] Tudor B, Várkonyi-Kóczy a R. Circular fuzzy neural network based hand gesture and posture modelling[C]// Instrumentation and Measurement Technology Conference. IEEE, 2010:815-820.

[6] Wang X, Xia M, Cai H, et al. Hidden-Markov-Models-Based Dynamic Hand Gesture Recognition[J]. Mathematical Problems in Engineering,2012, (2012-04-24), 2012, 2012(1):137-149.

[7] Wang X Y, Dai G Z, Zhang X W, et al. Recognition of Complex Dynamic Gesture Based on HMM-FNN Model[J]. Journal of Software, 2008, 19(9):2302-2312.