An Interactive Control System for 3D Display Using Hidden Markov Model

Liujing Xu[1], Xun Cao[2], Xiangdong Deng[3], Yuanqing Wang[4], Jing Jin

Abstract. This paper describes a system fulfilling interactive 3D-TV control. The system captures the position of the users' right hand, and then records and feature-extracts the trajectory of the gestures. After that, the movement of the users' input gestures can be recognized for interactive 3D-TV operations such as rotating a 3D object. In the system, the recognition of trajectory consists of three stages: the training of models, the pre-process of data and the estimation of gestures. Hidden Markov Model (HMM) with Left-Right Bounded (LRB) structure is used in the training and estimation while an improved direction coding is proposed to extract features. Besides, the methods of two-dimension filtering and corner detection are designed for the segmentation of gestures. The system can be applied to many applications (e.g. 3D game) with the gesture control section.

Keywords: Gesture Recognition, HMM, 3D display.

1 Introduction

 With the recent development of 3D stereoscopic display technology, 3D movies and 3D TV are becoming more and more popular in our daily lives. With the widely availability of the high-definition television (HDTV), several launched 3D broadcast TV channels enable us to experience 3D at home [1, 2].

---

[1] Liujing Xu, Jing Jin
School of Electronic Science and Engineering, Nanjing University, China

[2] Xun Cao(✉)
School of Electronic Science and Engineering, Nanjing University, China
e-mail: caoxun@nju.edu.cn

[3] Xiangdong Deng
The Academy of Broadcasting Planning, Beijing, China

[4] Yuanqing Wang(✉)
School of Electronic Science and Engineering, Nanjing University, China
e-mail: yqwang@nju.edu.cn

Currently, keyboard, mouse and touch-screen are frequently used as tools for human-machine interactions, which are not convenient enough. Directly using of hands as an input device is an attractive avenue for providing natural human-machine interactions [3, 4, 5].

Hidden Markov model (HMM) has become very popular to deal with the problems of gesture recognition because of its accuracy and flexibility. In HMM framework, the prior knowledge can be perfectly embedded in the Bayesian inter-ference, which can also be obtained by pre-trained data. The parameters of the HMM model can be estimated from a limited sample set with feasible training algorithm and the size and type of the model can be adjusted to meet the specific needs [6, 7]. We use a Left-Right Banded structure to perform HMM-based inter-active control of a 3D object and multiple initial methods to initialize the model parameters. The input gesture is described as a spatio-temporal sequence of fea-ture vectors.

In order to get better and seamless interaction, a wide variety of gestures are urgently needed in the system. However, as the quantity of the gestures increases, the number of models corresponding to each gesture will also increase. It is obviously a great burden in the training of models. On the other hand, the probabilities corresponding to each model in the identification and detection process have to be calculated for each gesture. That is also a big challenge for the real-time gesture recognition.

In this paper, a method of partial recognition and coding is proposed. A limited number of simple gesture models are selected, and then be used to construct recognizable gestures in certain order. In that way, the gesture to be recognized is only divided into several segments and the probability of every segment would be calculated separately. Then, it could be easily recognized by decoding. This method greatly releases the burden of computation to figure out the probability of the target gestures (gestures to be recognized) corresponding to each model. Likewise, the training of the model is also greatly simplified.
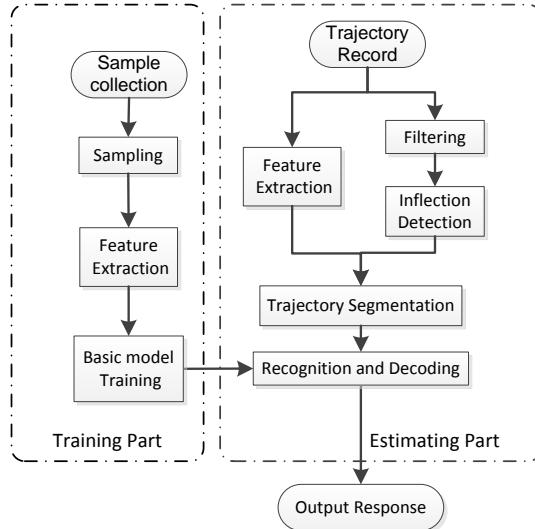
At the same time, an improved 8-direction feature vector is used to quantify characteristics to improve the detection rate, considering the users' operating habits. Besides, two-dimension filtering and corner detection are designed for this system specially.

With all the contributions above, we designed an interactive-control system for 3D-TV which collects gesture data from Kinect, using the trained model to recognize the implied meaning of the input gesture. The recognition result is verified by the rotation of a 3D cube.

2 Description of System

From the perspective of gesture recognition, our system consists of two parts: the training part using selected basic model samples and the estimating part using users' action sequences. HMMs of basic gesture models (short for Basic Model) are produced in the former part after some preprocessing. Then, users' actual ges-

ture in the latter part is divided into multi-segment sub-gestures according to the models and recognized separately. The diagram of the system is showed in Fig. 1.
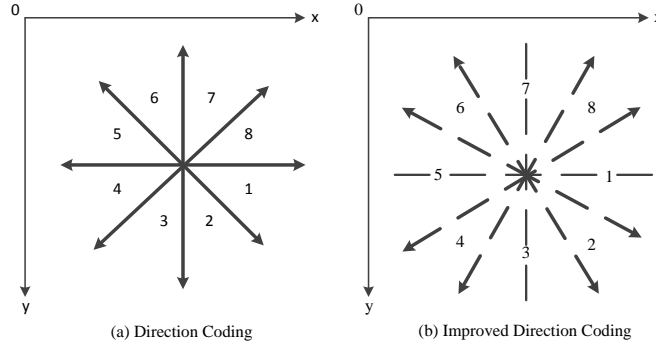

Fig. 1 System diagram

2.1 Collection of Experimental Data

The data used in this paper are collected by KINECT X-BOX 360. In the system, the images captured from kinectSensor are got in skeleton model through WindowSdk[8, 9]. Thus, right-hand coordinates of the user could be acquired from outstream of the skeleton which is also the original input of recognition. A sample of coordinate acquisition is showed in Fig. 2.

Fig. 2 Acquisition of right-handed coordinate

2.2 Feature Extraction of Sequences

(a) Direction Coding   (b) Improved Direction Coding

Fig. 3 Direction coding and improved direction coding

A direction primitive refers to one of n chain-code directions which are numbered 1 to n successively. In gesture recognition system, it represents the gesture direction from one point to the next, after quantization and code into one of n(called Direction Coding). Usually the area of every part is distributed equally and an 8-direction coding is showed in (a) of Fig.3 [10,11].

In this system, the sizes of the area for all directions are adjusted to fit the characteristics of operating habits. That is, the first, third, fifth and seventh areas are revised to $\pi/3$ and others $\pi/6$(called Improved Direction Coding and showed in (b) of Fig.3). Then the training samples and the gesture sequences could be converted to the corresponding coding sequences.

2.3 Hidden Markov Model

Hidden Markov Model is a double stochastic process consisting of an implicit and an explicit process [6, 7]. For model training and probability calculating, the Left-Right Bounded structure allows only self-transitions and transitions to the next indexed state. It is applied in our system with five states.

With multiple parameters initialized, the models are generated with the number being the same as the number of observation sequences [12]. We use this method for training the HMMs so that each observation sequence could be trained on its own initial model.

2.4 Basic Models and Codec Recognition

All target gestures designed in our system are composed of multiple sections with basic models. For better identification, the rightward model, upward model, leftward model and downward model are selected and trained correspondingly which are called Basic Model 1 to 4 in turn [13]. Thus, actual recognizable gestures of users are designed to consist of the four in a different manner.

Tab. 1 Accuracy rate of recognition based on Direction Coding

Model No.	1	2	3	4	others
Rate	96.6%	93.7%	98.9%	96.1%	98.2%

Tab. 2 Accuracy rate of recognition based on Improved Direction Coding

Model No.	1	2	3	4	others
Rate	97.0%	94.9%	99.4%	97.9%	99.0%

In the actual detection process, users' gestures are cut up into several segmentations. And each segment(called sub-gesture) will be separately recognized by the basic models and then got appropriate recognition model number. Finally, by decoding the recognition numbers of connected sub-gestures, the output terminal will make the appropriate response.

3 Test and Comparison of Method

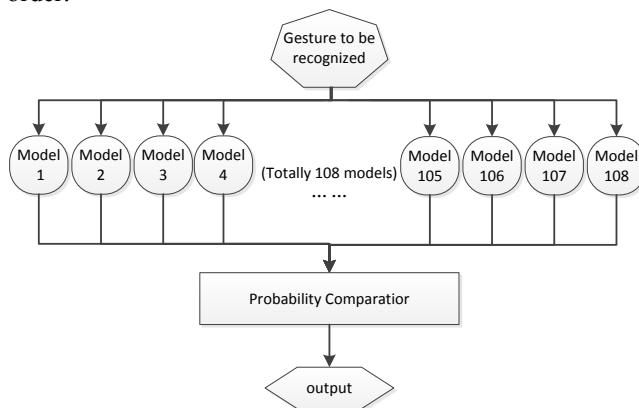### 3.1 Direction Coding and Improved Direction Coding

One hundred and twenty sets of samples each for four basic models are captured. Among them, twenty-five sets each are used for the training of every model and ninety-five sets each testing. The testing results for detection rate of Direction Coding and Improved Direction Coding are showed Tab.1 and Tab.2.

Compared with the traditional Direction Coding, Improved Direction Coding gets higher accuracy rates for models. Obviously, it has a more prominent performance in recognition.
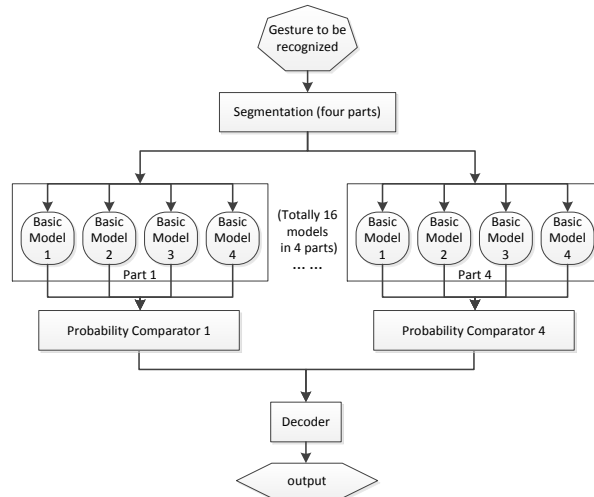
3.2 Ordinary Recognition and Partial Recognition

For a complete interactive system, a lot of gestures are required. It makes the whole system face a big problem in the model training and real-time detection.

Here, a limited number of simple basic gestures are determined and all recognizable gestures designed are combinations of the basic gestures in certain order.

Fig. 4 Ordinary recognition based on 108 models

Fig. 5 Partial recognition based on 4 basic models

For example, there are one hundred and eight different gestures required in a system. As a result, the same numbers of trained models are needed for the procedure of ordinary recognition. That is to say, one hundred and eight model-computations are done for one gesture's recognition. While in partial recognition method, only sixteen model-computations and some limited times of decoding computations are needed. Schematic diagrams of the two are showed in Fig. 4 and 5. Obviously, compared with ordinary recognition method, partial recognition is more efficient and consumes less time.
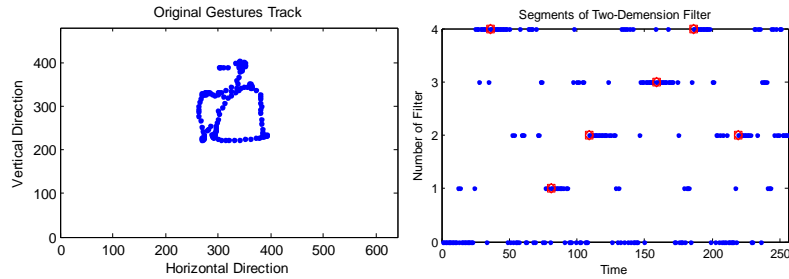
4 System Realization

A 3D cube with different textures on different surfaces is made as an instance to show the interactive result of recognition. It would rotate smoothly according to the basic gestures in (a) of Fig. 7. Here, a gesture of a rectangle with the start and end movements is taken into account as an example showed in (a) of Fig. 6.

The signal obtained by Kinect contains a certain degree of disturbance. Considering the characteristics of the models, two-dimension filter is used. That is, consider vector relationships between two adjacent points of a sequence and get the angle of each vector. Then, the reference numeral of each point could be obtained in accordance with the Tab. 3.

After two-dimensional filter, the gesture sequence is coded by the number 0 to 4. If the number of points has changed and remained for next six points, the sixth point is set as a corner. Finally, with starting and ending points added, all the corners can be obtained to cut up the gesture into several segments.

Tab. 3 Number of two-dimension filter

Range of α	(0,45)&(315,360)	(225,315)	(135,225)	(45,135)	Others
No.	1	2	3	4	0

Fig. 6 A gesture and its corner detection: (a) The trajectory of gesture (b) Two-dimension filter and corners
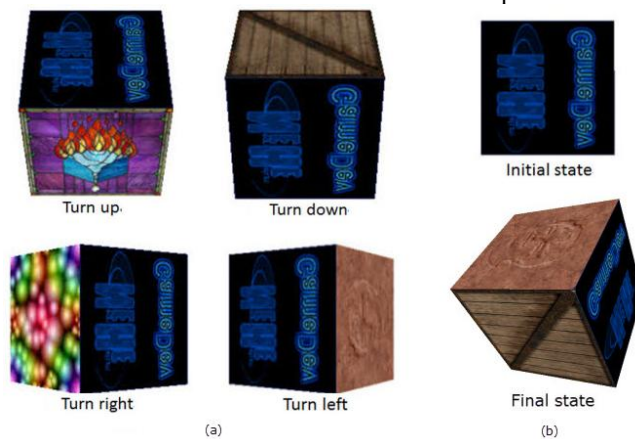
The result of the two-dimensional filter and corner detection for the gesture is shown in (b) of Fig. 6 (points with red square on them are the corners).

After the segmentations, the gestures are recognized separately by four basic models. Then, by comparing output probabilities of each model, the recognition number of every segment could be obtained. At the same time, the movement of the user's hand is recognized by decoding. And the final state of the 3D cube on the screen is shown in (b) of Fig. 7.

5 Conclusion and Future Works

The system described in this paper demonstrates the entire process of the collection and preprocess of the input gesture, the extraction of the feature, the training and recognition of models and eventually controlling the cube's turning based on the result of recognition. It is also a typical application of Human-Computer Interaction and can be used in an interactive 3-D display.

However, there are more future works to be done to improve the system. Firstly, the basic models can be extended to be compatible with more diverse gestures.

Fig. 7 A 3D-cube example of system: (a) Four corresponding turning of cube (b) The initial state and result of system

Secondly, the initial parameter can be further optimized to train more stable HMMs with faster speed of convergence and without local maximum points.

6 Reference

1. H. Yuan, J. Ćalić, and A. Kondoz, "Analysis of user requirements in interactive 3D video systems," Advances in Human-Computer Interaction, vol. 2012, 11 pages.
2. H.-T. Quan, M. Barkowsky, and P.L. Callet, "The importance of visual attention in improving the 3D-TV viewing experience: overview and new perspectives," IEEE Transactions on Broadcasting, vol. 57, no. 2, pp. 421-431, Jun. 2011.
3. P. Garg, N. Aggarwal, and S. Sofat, "Vision based hand gesture recognition," World Academy of Science, Engineering and Technology 49, pp. 972-977, 2009.
4. D. Ionescu, B. Ionescu, C. Gadea, and S. Islam, "An intelligent gesture interface for controlling TV sets and set-top boxes," 6th IEEE International Symposium on Applied Computational Intelligence and Informatics Timi oara, Romania, pp. 159-164, May 2011.
5. J. Shotton, A. Fitzgibbon, M.C.T. Sharp, M. Finocchio, R.M.A. Kipman, and A. Blake, "Real-time human pose recognition in parts from single depth images," Computer Vision and Pattern Recognition (CVPR), IEEE Conference, pp. 1297-1304, 2011.
6. L. Rabiner and B. Juang,"An introduction to Hidden Markov models," ASSP Magazine, IEEE, vol. 3, pp. 4 - 16, Jan. 1986.
7. S.R. Eddy, "What is a Hidden Markov model?" Nature Biotechnology, vol. 22, pp. 1315-1316, Oct. 2004.
8. "Programming Guide--Getting Started with the Kinect for Windows SDK Beta from Microsoft Research", Beta 1 Draft Version 1.0b, July 6, 2011.
9. "SkeletalViewer Walkthrough: C++ and C# -- Capturing Data with the NUI API", Beta 1 Draft Version 1.0a, June 24, 2011.
10. X.L. Li and D.-Y. Yeung,"On-line handwritten alphanumeric character recognition using dominant points in strokes," Pattern Recognition, vol. 30, no. 1, pp. 31-44, 1997.
11. X.Y. Wang, M. Xia, H.W. Cai, Y. Gao, and C. Cattani, "Hidden Markov Models Based Dynamic Hand Gesture Recognition", Mathematical Problems in Engineering, Vol. 2012.
12. N.J. Liu, R.I.A. Davis, B.C. Lovell, and P.J. Kootsookos, "Effect of initial HMM choices in multiple sequence training for gesture recognition," International Conference on Coding and Computing, vol. 1, pp. 608 - 613, 2004.
13. C.Y. Kao and C.S. Fahn, "A human-machine Interaction Technique: Hand Gesture Recognition Based on Hidden Markov Models with Trajectory of Hand Motion", Procedia Engineering, vol.15 pp.3739 – 3743, 2011.