Real-time Gesture Recognition for the Control of Avatar

ChanSu LEE, JiHyung LEE, ChanJong PARK, DongHyun KIM

VR Lab., Sensitivity Engineering Department, Systems Engineering Research Institute
(305-333 Ueun-dong Yusong-gu Taejun Korea, {chanse, jimmy, cjpark, dhkim}@seri.re.kr)

Abstract: There are many attempts to use the gestures for the natural human computer interface. Most of systems use the tracking devices to capture the human motion and display. In this paper, we propose a gesture recognition system for natural interface between the participant and his own avatar in virtual environment only using two polhemus sensors attached on the arm. Our system has two modes for the control of avatar. One is tracking gestures. In this mode, tracking data is used only to simulating the motion of the participant's one arm. The other is the interpretation of gestures. In the interpretation mode, the participant can make his avatar walk or run by the speed corresponding to the velocity of gesture and command his avatar to work properly by natural gesture. This gesture interpretation is done in real-time and may give an immediate response in virtual environments. Therefore, our system provides more natural interface in virtual environments by the gesture.

***Key Words**: Gesture Recognition, Avatar Control, Interface, Fuzzy Rule*

1. Introduction

Perhaps human hand is the most useful tool to interact with the environment we live. We manipulate the physical world most open and most naturally with our hand. Along with speech and facial expression, hand gestures are used as one of the primary methods of communication between people [1]. Hand gesture is also useful to communicate between the deaf. And hand gesture recognition systems to interpret sign language for the deaf have been developed [2-4].

Virtual environment is the representation of a computer model or database, which can be interactively experienced by the virtual environment participant(s) [5]. Nowadays there are many attempts for effective interaction with the virtual environment [6].

One of the attempts to make the effective interface with virtual environment is using gestures. A key reason for using hand gestures in human-computer interaction is to take advantage of natural and intuitive communication methods [7]. We use hands to manipulate goods and to communicate symbolic meanings. But most of the virtual reality systems only use hand device to manipulate virtual objects in virtual environment by tracking the hand gesture. Some of them only recognize the hand gesture and represent symbolic meanings. It is necessary to use human hand gesture not only for manipulating virtual environment by sensing data but also for symbolic communication with virtual environment to be more useful interface.

An avatar is the representation of participant in virtual environment. The avatar is required to have active autonomous behaviors together with a series of strings which the user is continuously pulling as smoothly as possible [8]. Using the avatar, participant can be represented easily and interact with other participants in virtual environment.

This paper propose a hand gesture recognition system, especially arm gesture, which recognize the pre-defined symbolic meaning for the control of avatar's basic motion as well as tracking motion of the participant for the avatar slaving.

The paper is organized as follows. In section 2, overall system structure is described. And gesture for the control of avatar is defined. Direction recognition classifier and orientation recognition classifier are explained in section 3. In section 4, shows the primary result. And then follow conclusion and acknowledgement.

2. Overview of Gesture Recognition System

Most hand gesture recognition system is based on glove devices or vision systems [8]. Glove devices are more accurate and fast than vision-based systems. But it's more expensive and encumbered than vision systems. In case of vision based system, it is difficult to calculate the position of hand or arm and difficulty in real time processing.

For the Virtual Office Environment Systems (VOES), which can have multi-participants, both of them are not proper for expensive or difficulty in real-time processing. This gesture recognition system uses the arm data for the control of the avatar. Using this arm data, we can control the avatar by slaving the motion of the participant's one arm. Also we can command the avatar by the recognition of the arm gesture. We attached two polhemus sensors on the forearm and upper arm.

For the recognition of the hand gesture, at the first position and orientation data for the forearm and upper arm are sensed. The data are transformed for the proper coordinate system according to virtual environment.

The position data is used to estimate motion states. These motion states decide current state of the gesture considering the previous movement of gesture. It is also used to determine whether the given gesture is meaningful or not by checking the state flow.

Proper functions are called to recognize the basic elements of the gesture for each state. After that, these recognition results for basic elements are interpreted to recognize the meaning of gestures. We have two interpretation modes. The first mode is just to track data used in simulating the motion of the participant's. By which data, we can generate more realistic motion of avatar. The second mode is interpretation of gestures as commands. By gesture command, we may control avatar with symbolically.

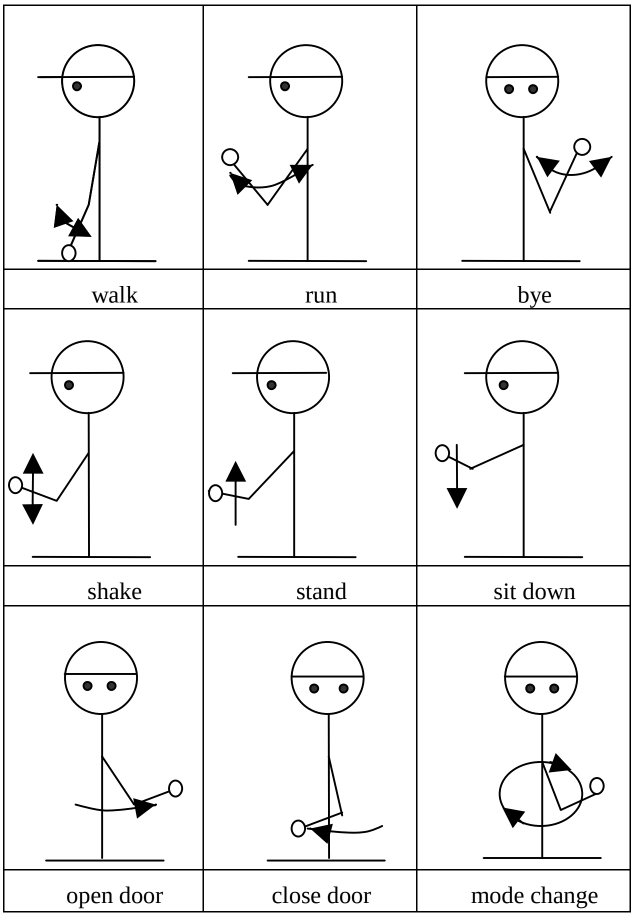

Fig. 1 The defined command gestures

For the control of the avatar, we define 9 command gestures. Fig. 1 shows defined gestures. 8 of them are commands to control the avatar. Last one is the command to change mode. 8 command gestures are 'walk', 'run', 'bye', 'shake', 'stand', 'sit down', 'open door', and 'close door'. These are come from basic motion for generation of avatar control. In cases of walk and run, we calculate the velocity of the gesture to move avatar according to the moving velocity of the participant, which may give direct and easy method for the control of avatar according to user's intention. By using these techniques, we can easily control the motion of the avatar.

In mode changing gesture, we make a circle, which is easy and can be drawn with the same start point and end point. At first, it changes the mode to command mode. In command mode, this gesture changes mode to tracking mode. In tracking mode, this gesture changes mode to command mode.

3. Gesture Recognition

The system structure for the gesture recognition is composed of 4 stages. Fig. 2 shows the system configuration.

The first stage is data sensing for forearm and upper arm and coordinate transformation. With polhemus sensor, this system get a pair of 3D position data (x, y, z) and rotation data (roll, pitch, yaw) for the forearm and upper arm. Coordinate transformation is done according to the virtual environment.

The second stage is the state estimation. According to the velocity of motion and change of velocity, we partitioned 10 different state which are similar to state for Korean sign language recognition system [4]. For the state estimation, we use motion partition and finite state automata (FSA).

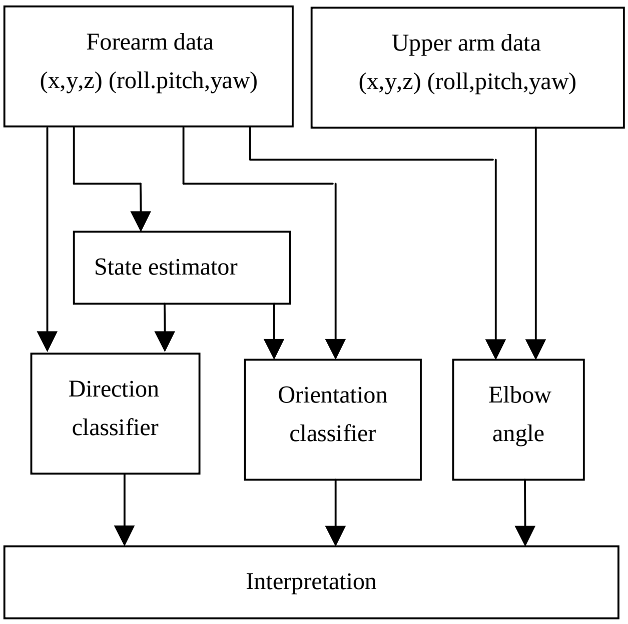

Fig. 2 System configuration

The third stage is the basic element recognition for each gesture by Fuzzy and Neural Networks classifier. For the recognition of gesture, we use 2 kinds of classifiers. One is the direction classifier by fuzzy rule. The other is the orientation classifier by Fuzzy Min-Max Neural Networks (FMMNN) [9]. These classifiers are executed according to the state. Using this classifier, we recognize the basic elements of gestures.

The fourth stage is to interpret the meaning of gesture and send command or to track arm data to simulate of the avatar.

The interpretation of gesture is done considering the position classifier results, the orientation classifier results and an angle of the elbow.

3.1. Direction Recognition

Direction classifier recognizes 8 basic direction elements. As there are time-scale variance, dimensional complexity and no staring/ ending point in hand gesture, the recognition of dynamic hand gesture is difficult [10]. To overcome such difficulties, this system uses the state automata to segment the gesture and to find the beginning and ending of gesture. To overcome the problem of time scale variance, this system use feature extraction method [11] for recognition of each gesture. And to overcome variation of gesture in each person and to get the accuracy of the executed gesture, we used fuzzy rule.

Features are selected to distinguish 8 basic elements of direction effectively. The selected features are as follows. ‘k’ is current sampling time.

(a) Relative end position (EP): relative end position of gesture from the beginning point of the gesture.

$$EP_{x,y,z} = R_{x,y,z}(k) - R_{x,y,z}(0) \qquad (1)$$
$$R_{x,y,z}(0): initial\ position\ of\ gesture$$
$$R_{x,y,z}(k): k'th\ position\ data\ for\ each\ axis$$

(b) Cumulative length of motion (CL): length of motion trajectory for each axis.

$$CL_{x,y,z}(k) = CL_{x,y,z}(k-1) + L_{x,y,z} \qquad (2)$$
$$L_{x,y,z} = \left| R_{x,y,z}(k) - R_{x,y,z}(k-1) \right|$$

(c) Total direction change (TC): sum of direction change in each sampling time.

$$TC_{xy}(k) = TC_{xy}(k-1) + \left(\tan 2^{-1}\left(\frac{R_y(k)-R_y(k-1)}{R_x(k)-R_x(k-1)}\right) - \tan 2^{-1}\left(\frac{R_y(k-1)-R_y(k-2)}{R_x(k-1)-R_x(k-2)}\right) \right) \qquad (3)$$

(d) Change of Direction: direction change for each axis.

$$CD_x = CD_x + f(x)$$
$$CD_y = CD_y + f(y) \qquad (4)$$

$$f(x) = \begin{cases} 1 & if\,[R_x(k) - R_x(k-1)] \times [R_x(k-1) - R_x(k-2)] < 0 \\ 0 & otherwise \end{cases} \qquad (5)$$

Using this features and fuzzy rule, this system classifies 8 different direction class.

3.2. Orientation Recognition

Polhemus senses rotation data as Euler angle. Using this data, this system classifies orientation of each gesture at the beginning. We use Fuzzy Min-Max Neural Networks [9] for the recognition of orientation. FMMNN can have many hyper-boxes for one class. And, the membership function of the hyper-box is expressed as follows.

$$f(A_h, V_j, W_j) = \frac{1}{N}\sum_{i=1}^{N}[1 - f(a_{hi} - w_{ji}, \gamma) - f(v_{ij} - a_{hi}, \gamma)] \qquad (6)$$

$$f(x,\gamma) = \begin{cases} 1 & , x\gamma > 1 \\ x\gamma & , 0 \le x\gamma \le 1 \\ 0 & , x\gamma < 0 \end{cases}$$

In equation (6), **A** is input vector for rotation data, which is normalized to [0,1]. And, **V, W** define min and max points of fuzzy set hyper-box. N is number of input dimension.

For orientation classifier, initial min-max values of the networks are determined by training with sample data. 5 basic elements are defined to identify similar direction gesture.

4. Experiments

We run this system in Silicon Graphics Computer, Octane and 2 Fastrak, for the tracking sensors. Fig. 3 shows initial experiment results. we made gesture more than 20 times for each gesture.

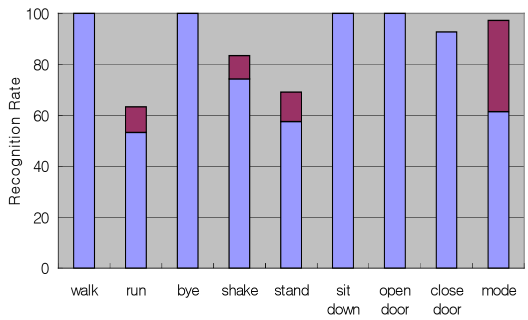

Fig. 3 Gesture Recognition Rate

Recognition rate is 82.15% in average. To improve this gesture recognition system, we use orientation limitation in each gesture interpretation and improved the recognition rate. Still 'run' gesture recognition rate is not good. So we use elbow angle. Using this we can improve the system.

In addition to this basic command gesture, this system calculates the velocity of motion in case of walk and run to control avatar more easily.

5. Conclusion

To make useful interface to control of avatar, we show hand gesture recognition system. Using this system, we can command avatar and simulate participants motion. This system also shows effectiveness of fuzzy rule and fuzzy min-max neural networks for the recognition of gesture basic elements.

6. Acknowledgement

The part of this work is supported by Ministry of Information and Communication (MIC).

References

[1] J. S. Kim, C. S. Lee, Z. Bien, B. T. Choi, and K. J. Song: A Dynamic Hand Gesture Recognition System for Virtual Reality, IFSA'97, 1997.

[2] T. Starner and A. Pentland: Visual Recognition of American Sign Language using Hidden Markov Models, Int. Workshop on Automatic Face- and Gesture-Recognition, pp. 189-194, June 1995.

[3] M. S. Waldron and S. Kim: Isolated ASL Sign Recognition System for Deaf Persons, IEEE Trans. Rehabilitation Engineering, vol. 3, pp. 261-271, Sep. 1995.

[4] C. S. Lee, G. T. Park, J. S. Kim, Z. Bien, W. Zaong, and S. K. Kim: Real-time Recognition System of Korean Sign Language based on Elementary Components, FUZZ-IEEE'97, pp. 1463- 468, July 1997.

[5] T. A. Furness and W. Barfield: Introduction to Virtual Environments and Advanced Interface Design, Oxford University Press, 1995.

[6] A. P. Pentland: Smart Rooms, Scientific America, pp. 54-62,April 1996.

[7] K. H. Quek: Unencumbered Gesture Interaction, IEEE Multimedia, pp. 36-pp.47, 1996.

[8] S. Benford, J. Bowers, L. E. Fahlen, C. Greenhalgh, and D. Snowdon: Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds, Multimedia Systems, Springer-Verlag, pp. 93-104, 1997.

[9] P. Simpson: Fuzzy Min-Max Neural Networks-part 1-Classification, IEEE Trans. on Neural Networks, Vol. 3, pp. 776 – 786, Sep. 1992.

[10] Y.H. Nam and K. W. Wohn: Recognition of Space-Time Hand-Gestures using Hidden Markov Model, VRAIS'95, 1995.

[11]D. Rubine: Specifying Gestures by Example, SIGGRAPH'91 conference Proceedings, pp. 329-337, Aug. 1991.