Body Gesture Recognition Framework for 3D Interactive Systems

Choonsung Shin, Jisoo Hong, Yougmin Kim, Sung-Hee Hong, Hoonjong Kang
Realistic Media Platform Research Center
Korea Electronics Technologies Institute
Seoul, Republic of Korea
e-mail: {cshin, jhong, rainmaker, shhong, hoonjongkang}@keti.re.kr

***Abstract*—This paper proposes a body gesture recognition framework for 3D interactive systems. For this purpose, we built the gesture recognition framework via sampling, pre-processing, quantification, and building an inference model using the hidden Markov model (HMM). Using this framework, gestures are detected and classified from 3D trajectory data in real time. Evaluation proved that our recognition system successfully supports 3D interactive systems. We also tested the proposed framework using 3D virtual training systems.**

Keywords- gesture recognition; 3D interaction.

I. INTRODUCTION

With the recent advance of 3D technologies, a number of studies have been proposed to understand and recognize users' positions and gestures. In a virtual environment, such information is very important in presenting a 3D immersive environment [1]. The 3D locations and gestures of the hands and a head are especially important in allowing users to interact naturally with the 3D environment and are widely used for interactive applications.

For this purpose, we proposed a 3D gesture recognition framework for 3D interactive applications. The proposed framework consists of 3D gesture recognition and inference components. The recognition component collects and preprocesses the 3D locations of the head and hands then extracts features and builds models for inferring gestures. We applied the hidden Markov model (HMM) for modeling and inferring gestures since the gestures are made in temporal sequence [2]. The inference component then detects and selects the target gesture from the 3D trajectory of body by inferencing probability of gesture candidates in real-time. The resulting gesture and position information are delivered to 3D interactive systems.

The remaining part of this paper is organized as follows. We first define target gestures for 3D interactive systems and then introduce the gesture recognition framework in Section II. We then show how we implemented our framework in Section III. Finally we conclude with future work in Section IV.

II. RECOGNIZING BODY GESTURES

A. *Target Body Gestures*

First we define the body's gestures that will be used for 3D interactive systems. We are interested in interactive systems where a user is set at a fixed position in front of the system. Thus we need to detect gestures from different parts of the body. We selected upper body gestures, hand gestures, and head gestures (see Figure 1). The head gestures include left, right, up, and down movements. The hand gestures are represented as grabbing, pointing, and releasing. The upper body gestures include left, right, up, down, and directional pointing gestures.

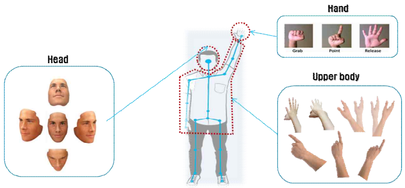

Figure 1. Target Body Gestures.

B. *Body Gesture Recognition Framework*

In order to recognize the gestures defined, we built a gesture recognition framework. Previous work has only focused on modeling and inferencing gestures [3]. Here we added more steps, such as gesture sampling, segmentation and labeling, pre-processing, quantification, and establishing an inference model. Figure 2 shows the overall procedure of building our gesture recognition framework.

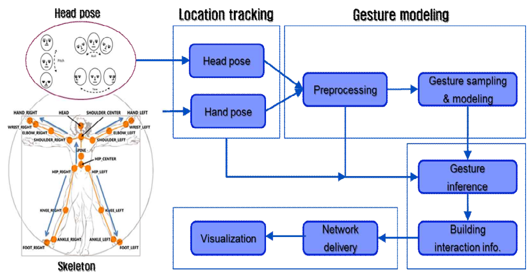

Figure 2. Overall Procedure of Body Gestures Modeling.

In the sampling, the framework collects the raw position data from a location tracking sensor. The sensor detects the body and generates the 3D position of the body parts. This sampling phase collects the raw data of all the body parts and extracts the segment of each body gesture. In the pre-processing phase, the noise is removed by filters, such as moving average, and the relative movement value is extracted for normalization. Afterward we reduce data dimension and extract features by using the *k*-means algorithm. Finally, we built an inference model with the

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-468-8

features selected. As a body gesture consists of a temporal sequence, continuous machine learning algorithms are useful. We especially selected HMM since it is robust on individual differences of gestures and small numbers of samples.

C. *Inferencing Body Gestures*

The inference phase consists of sampling, pre-processing, detection, quantification, and inference. Thus the sampling, pre-processing, and quantification are the same as those of the modeling, but the detection and inferences are additionally required. In order to detect the existence of a gesture from 3D trajectory data, we assume that there is a small amount of 3D movement when a user is in the ready state. For this purpose, we apply a threshold to the 3D movement during a certain period of time called a window. The window ranges from 0.5 s to 2 s and is related to the length of a gesture. The amount of 3D movement is represented by the sum of the standard deviation of each axis. The threshold is less than the maximum value of the sum of the standard deviation of the gesture samples. When there is a gesture in the 3D trajectory, the quantification is applied to the trajectory data and the probability of each gesture is calculated from the inference model. Finally the gesture with the highest probability is selected as the result of the inference.

III. EVALUATION

We implemented it in a Windows 7 environment. In order to obtain fast data, we used a Microsoft Kinect version 2 sensor which provides a 1080p RGB image stream with a 640 x 480 depth map [4]. We then collected the 3D locations of the heads, hands, and upper bodies of study participants and modeled seven gestures made from hand trajectories. We selected eight upper body gestures: seven gestures (Nos. 2-8) and one ready statue (No. 1). We collected 203 gesture samples from three participants. We evaluated our framework based on a 10-fold cross validation.

As seen in Figure 3, the accuracy of the proposed framework is about 96% when three hidden nodes were used in the framework. The accuracy reached 98% when five hidden nodes were used. The accuracy was not improved even when more hidden nodes were used.

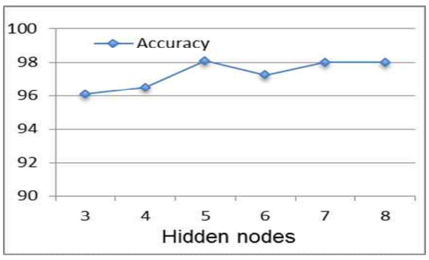

Figure 3. Accuracy of Body Gesture Recognition.

We also analyzed the quality of the individual gesture classification. We calculated the confusion matrix of the gesture recognition classification. As seen in Figure 4, the main errors were from the ready status. Three of the 25 ready status samples were misclassified as Gesture 2. Other gestures were correctly classified.

Gesture	1	2	3	4	5	6	7	8
1	**22**	**3**	0	0	0	0	0	0
2	0	**25**	0	0	0	0	0	0
3	0	0	**28**	0	0	0	0	0
4	0	0	0	**27**	0	0	0	0
5	0	0	0	0	**27**	0	0	0
6	0	0	0	0	0	**29**	0	0
7	0	0	0	0	0	0	**10**	0
8	0	0	0	0	0	0	0	**33**

Figure 4. Confusion Matrix of Body Gesture Recognition.

We also tested the proposed framework by connecting with two virtual training systems, as illustrated in Figure 5. One of the systems was a virtual train training simulator and another one was a virtual driving simulator. In order to detect a gesture from the continuous 3D trajectory of a hand, we used a 2-second window with a 0.5 second overlap. The framework then selected the gesture with the highest probability inferred from HMM in real-time. Finally the position data and resulting gesture is delivered to the systems.

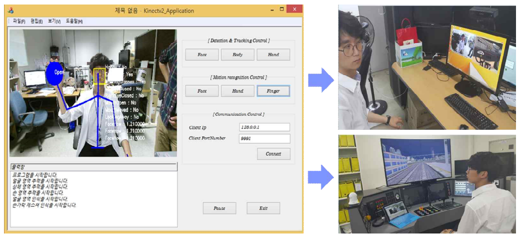

Figure 5. Testing with the Virtual Training System.

IV. CONCLUSION

This paper proposed a body gesture recognition framework for 3D interactive systems. For this purpose, the proposed framework collected, segmented, pre-processed, and quantified 3D trajectory data from a 3D tracking sensor and then built an inference model for recognizing body gestures. Later the framework detected and classified the gestures from the 3D trajectory data by inferencing their probabilities. We also evaluated the proposed framework and tested it with 3D interactive systems.

ACKNOWLEDGMENT

This paper is supported by the Ministry of Science, ICT, and Future Planning (MSIP) (Cross-Ministry Giga KOREA Project).

REFERENCES

[1] J. Weissmann and R. Salomon, "Gesture recognition for virtual reality applications using data gloves and neural networks," in Neural Networks, 1999. IJCNN '99. International Joint Conference on, Vol.3,1999, pp.2043-2046.

[2] Z. Ghahramani, "An Introduction to Hidden Markov Models and Bayesian Networks", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 15, No 1, 2001, pp 9-42.

[3] C. Shin and B. Park, "3D gesture modeling and recognition for Immersive Virtual Reality", DHIP 2015, pp 59.

Copyright (c) IARIA, 2016. ISBN: 978-1-61208-468-8