Eye Gaze, Hand Pointing and Head Movements Feature Based Human Intention State Recognition

Suparat Yeamkuan and Kosin Chamnongthai

***Abstract*— Eye gaze, hand pointing, and head generally approach the point in the same order when a human expresses an intention on a 3D point, and eye fixation during intention expression is a key to set the intention point position. This paper proposes a multimodal fusion of eye gazing, hand pointing, and head to determine the state of 3D intention. In order to reach the 3D point for intention expression, the method uses human psychovisual knowledge in the sequence of eye gaze, hand pointing, and head as a pattern to determine intention mode. In the intention state, the position of the intention point in three-dimensional space is primarily determined by the sequence of eye gaze, hand pointing, and head velocity.**

I. Introduction

One of the simplest methods for aged, disabled, and normal individuals to convey their intentions with a machine is to touch or push a switch button. Passengers in an auto-driving car without drivers can seat anywhere they want and act as they want in 3D space. Many switch buttons, which are vital tools for passengers or users to communicate with machines, may be located physically and virtually everywhere, causing switch button placements to grow distant from users. These truly need a handy feature that consumers may use whenever and wherever they choose. A function to automatically recognize and trigger virtual and physical switch buttons that are sometimes installed at a distance is required to activate virtual and physical switch buttons that are sometimes installed at a distance.

II. Proposed Method

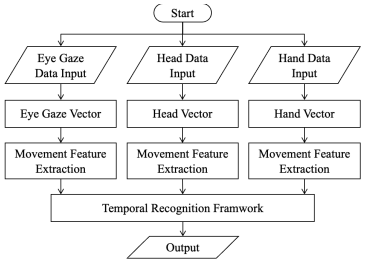

Fig.1. Flowchart of proposed method.

This paper introduces a method to identify from eye, head, and hand vector data, feature extraction and use of recognition framework in the context of temporal generally. The method proposed in the paper involves mathematical formulations that incorporate eye, head, and hand vectors along with movement features into the speed of movement. Subsequently, those formulations of the movement feature input requirements of the temporal recognition framework [1] as shown in Fig 1.

III. Experimental Results

To evaluation of the proposed recognition involved ten examiners, comprising eight males and two females. The experimental were allows each examiner afforded the flexibility to move their head within the limitations imposed by the eye tracker's distance constraints as shown in Fig 2.

Fig. 2. Examiner during perform the experiment.

TABLE I. Acuracy of Experiment

Technique	Accuracy
SVM	85.17%
RNN	88.73%
LSTM	89.81%

Conclusion

This paper introduces a method for recognizing intention states using eye, head, and hand data in time sequences, incorporating speed movement features. The temporal recognition with RNN and LSTM models, along with the speed movement feature giving, successfully identifies the intention state in human hand-pointing interactions. Based on approach utilizes the speed movement features of eye, head, and hand information, considering the temporal sequence, and demonstrates the potential of RNN and LSTM models in recognizing hand-pointing tasks through natural eye, head, and hand movements. The performance of the proposed method indicates a compromise in achieving intention state recognition for human-computer interaction tasks. The average accuracy of temporal recognition across all experimental tasks is 89.27% effectually.

References

[1] Sherstinsky, A. 2018. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) Network.

Suparat Yeamkuan is with the Faculty of Computer and Robotics Engineering, Faculty of Engineering, Bangkok University, Bangkok, Thailand. (e-mail: suparat_y@bu.ac.th).