4ZF-06

Analysis and Synthesis of Eye Movements for Conversational Agent

Bowen Hu Daichi Ando Tomohiko Mukai
Tokyo Metropolitan University

1. Introduction

3D animation of conversational agents becomes popular in our daily life, such as Samsung's Neon. Eye movement of the agent is sometimes more important non-verbal information than words to share what they think and how they feel in face-to-face conversation. Funda [1] presented a generative adversarial learning approach with captured time-series data to synthesize continuous gazing behavior. However, they do not consider behavioral mode in the conversation and do not explicitly model the two types of eye movements: saccades with relatively large movements and fixation with the minor movement over a longer time [2]. Our key insight is that the direction of eye movement and the frequency of movement will change depending on the behavioral mode, such as thinking and listening.

We propose a method to generate natural eye movements by switching states in finite automaton based on analysis of eye-tracking data. Fig. 1 shows the overall architecture of our system. First, we collect eye movement data on three modes: listening, thinking, and speaking, from real conversation. Next, a finite automaton for each behavioral mode is built based on the analysis of the eye movement data. Finally, the eye movement of a virtual agent is synthesized using the finite automaton and gaze point transformation model.

2. Proposed Method

Finite automaton is a mathematical model suitable for modeling the dynamic behavior of an object. In the analysis stage, two elements of the finite automaton are learned from the real data.

The first element, state, represents the similar gazing points group estimated using the K-means clustering method. Each cluster's center coordinate is used as the base point of eye movements. The number of clusters is manually determined according to the silhouette score, elbow method, and the actual clustering results.

The second element is the transition probability between the base points. The transition probability can be estimated from the acquired temporal sequence of cluster labels. To distinguish between the two types of eye movements in this part, we count it as a saccade when the gaze point stays in fixation too short and is accompanied by the next movement to another cluster.

In the runtime process, position of the gaze point is determined according to the state transition in the finite automaton. Then, eyes balls will automatically rotate based on the three-dimensional transformation. Finally, the runtime module generates eye movements of a virtual character.

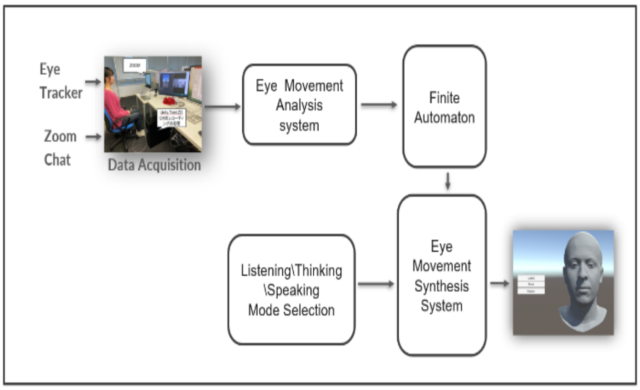

Fig. 1: Overall architecture

3. Experiment

Data acquisition setup is shown in the upper left part of Fig. 1. Eye-tracking with the Tobii 4c was performed on the Unity platform on a desktop computer. The participant was 24 years old native Chinese international student with normal corrected visual acuity. The experimental tasks were to design a route from destination to her home and recall her old experiences. Then we used data acquisition setups to record raw eye-movement data of the three modes in the screen space at 90 Hz.

In the analysis of thinking mode, the gaze points were clustered into five groups, as shown in Fig. 2. The resulting finite automaton model is shown in Fig. 3. The coordinate value in each state is the eye movement's base point. The arrows indicate the direction of eye movement, and the thickness of the arrows indicates the ratio of transition probabilities. With the above three elements, the finite automaton in the thinking mode shows the path of eye movements.

As a result, with the state transitions in the finite automaton, our method generates the realistic eye movements shown in Fig. 4. Two types of eye movement features, such as the duration of a saccade, that were not produced by previous techniques, were also generated using our method.

Fig. 2: Analysis of thinking mode using K-means clustering

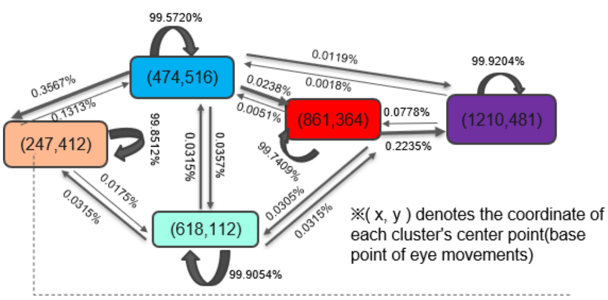

Fig. 3: Finite automaton of think mode

Fig. 4: Sample images of eye movement generation

4. Discussion

The proposed method synthesizes eye movements using finite automaton based on behavior mode. However, eye movements generated in this research did not consider the coordination of blink and head animations. Our future work includes using natural language processing to replace the manual switching of behavioral modes.

Acknowledgment

We would like to thank Meixuan Li, Youdi Li, and Yibin Yan for the eye-tracking data. This work was supported by JSPS KAKENHI Grant Number 19H04231 and 21K12193.

Reference

[1] Funda Durupinar, Personality-Driven Gaze Animation with Conditional Generative Adversarial Networks, The 13th Annual ACM SIGGRAPH Conference on Motion, Interaction and Games (MIG20), 2020.

[2] https://www.tobiipro.com/learn-and support , Types of eye movement. (Last access: December 21st, 2021).