Yunzhan ZHOU, Tian FENG, Shihui SHUAI, Xiangdong LI, Lingyun SUN, Henry Been-Lirn DUH, 2022. EDVAM: a 3D eye-tracking dataset for visual attention modeling in a virtual museum. *Frontiers of Information Technology & Electronic Engineering*, 23(1):101-112. https://doi.org/10.1631/FITEE.2000318

EDVAM: a 3D eye-tracking dataset for visual attention modeling in a virtual museum

Key words: Visual attention; Virtual museums; Eye-tracking datasets; Gaze detection; Deep learning

Corresponding author: Tian FENG
E-mail: t.feng@zju.edu.cn
ORCID: https://orcid.org/0000-0001-9691-3266

Motivation

1. Predicting visual attention facilitates an adaptive virtual museum environment and provides a context-aware and interactive user experience.
2. Explorations toward development of a visual attention mechanism using eye-tracking data have so far been limited to 2D cases, and researchers are yet to approach this topic in a 3D virtual environment and from a spatiotemporal perspective.
3. The related datasets are still limited to 2D cases and are not labeled regarding timestamps. They could hardly represent sequential behaviors or support context-aware interaction in a 3D virtual environment.

Main idea

1. Present a 3D eye-tracking dataset.
2. Propose a novel approach to achieve gaze-based 3D interaction.
3. Devise a deep learning model to predict a user's visual attention in the next moment from previous records.

Method

1. We present a 3D Eye-tracking Dataset for Visual Attention modeling in a virtual Museum, named EDVAM. EDVAM includes 9 604 480 visual attention records from users during their navigation. We divide these records into two subsets: the raw subset holds the captured eye movement sequences and the practical subset comprises the processed samples.

Table 1 Types of collected data

Type	Number of features	Frequency	Device
VR gaze	11	> 30 Hz	VR headset, eye-tracking gadget
Pupil	33	30 Hz	Eye-tracking gadget

Method (Cont'd)

2. To enable gaze-based 3D interaction with virtual objects, we propose a novel approach that maps 2D PoG positions to the corresponding 3D positions in VR, which contribute to obtaining real-time eye movements for recording visual attention.

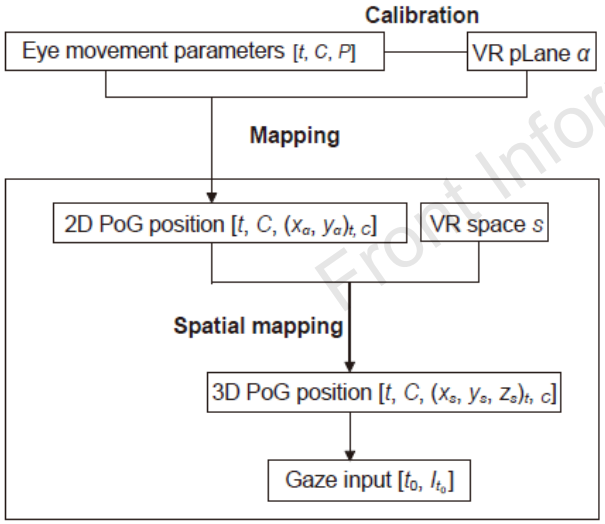

Fig. 2 Pipeline of gaze-based 3D interaction

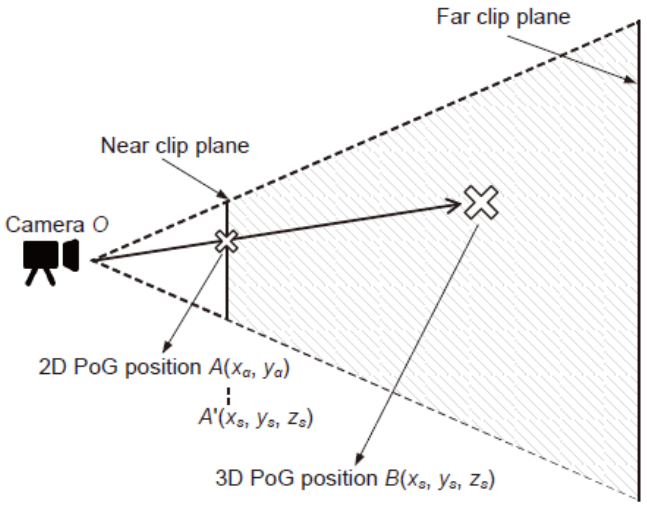

Fig. 3 Mapping a 2D PoG position to the corresponding 3D PoG position in the VR space

Method (Cont'd)

3. To predict the visual attention and validate the collected dataset, we devise a three-layer long short-term memory (LSTM) network deep learning model to predict a user's visual attention in the next moment from previous records.

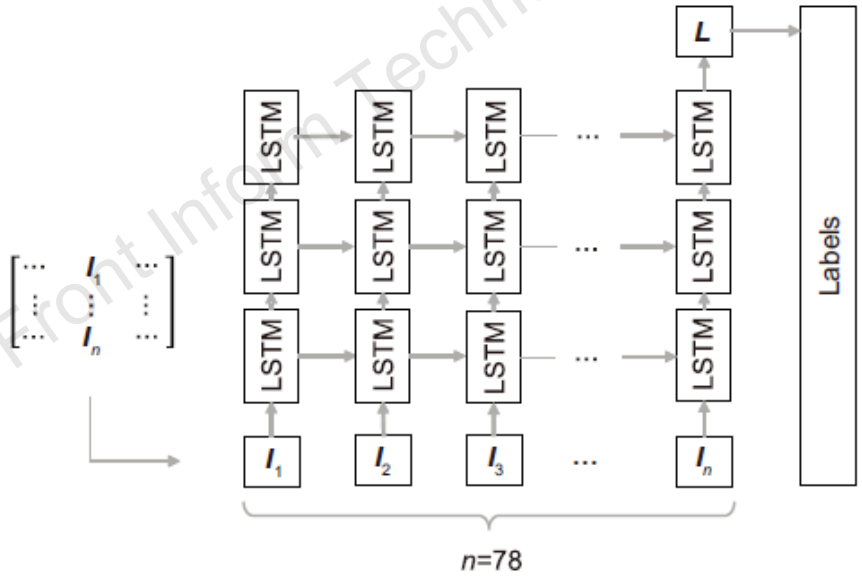

Fig. 5 Architecture of the three-layer LSTM network

Major results

The cross-entropy loss over 40 epochs

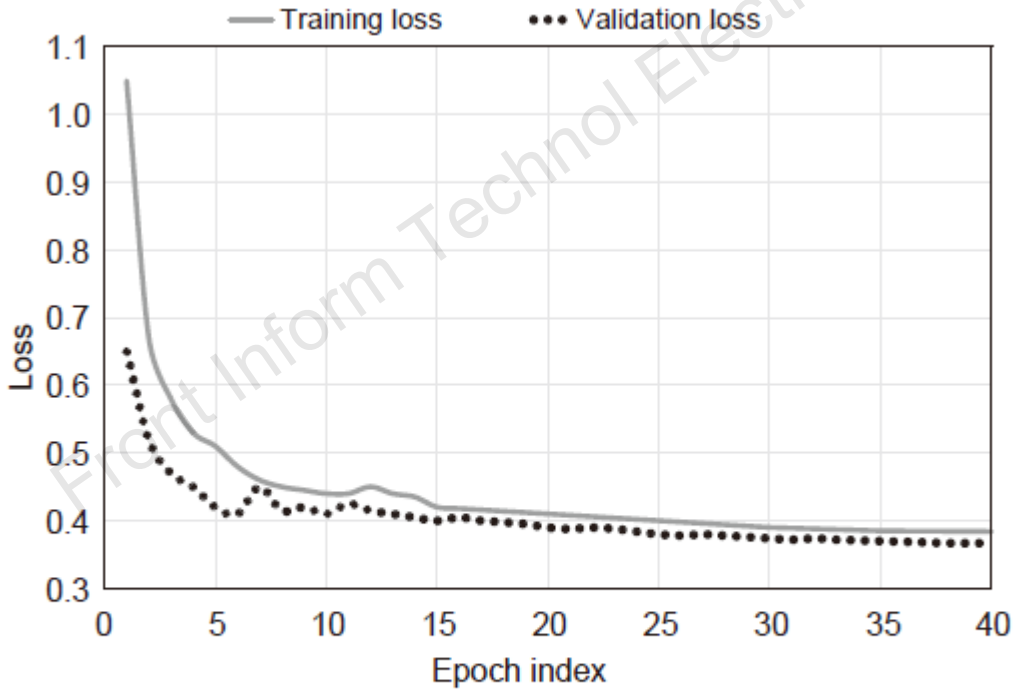

Fig. 6 Per-epoch loss when training and validating the LSTM network

Major results (Cont'd)

Prediction example

Fig. 7 An example of predicting the next visual area: (a) 3D PoG position sequences (white trajectories) and the camera's position sequences (black trajectory) visualized inside the virtual museum; (b) the camera's orientation sequences represented by the user's head motions; (c) the next visual area predicted by the model; (d) the corresponding ground truth

Major results (Cont'd)

Analysis of gender effects

Table 5 Prediction accuracy of the model regarding the user's gender

Model case	Accuracy (%)	
	Female	Male
As in Section 4.3	79.80	79.87
Trained on data from female users	77.99	77.15
Trained on data from male users	77.88	77.02

Conclusions

1. We introduced EDVAM, the first 3D eye-tracking dataset in a virtual museum, and proposed a predictive model for visual attention based on previous eye movements. Our model, based on the LSTM network, supports fundamental context-aware interactions in a 3D virtual museum.

2. Our work contributes to enabling a virtual museum's adaptiveness for a context-aware user experience. It helps users interact with virtual objects and adaptive UIs through a personalized virtual museum tour.