
IEEE
VR
IEEE
2022
THE 29TH IEEE CONFERENCE ON VIRTUAL REALITY & 3D USER INTERFACES

EHTask: Recognizing User Tasks from Eye and Head Movements in Immersive Virtual Reality

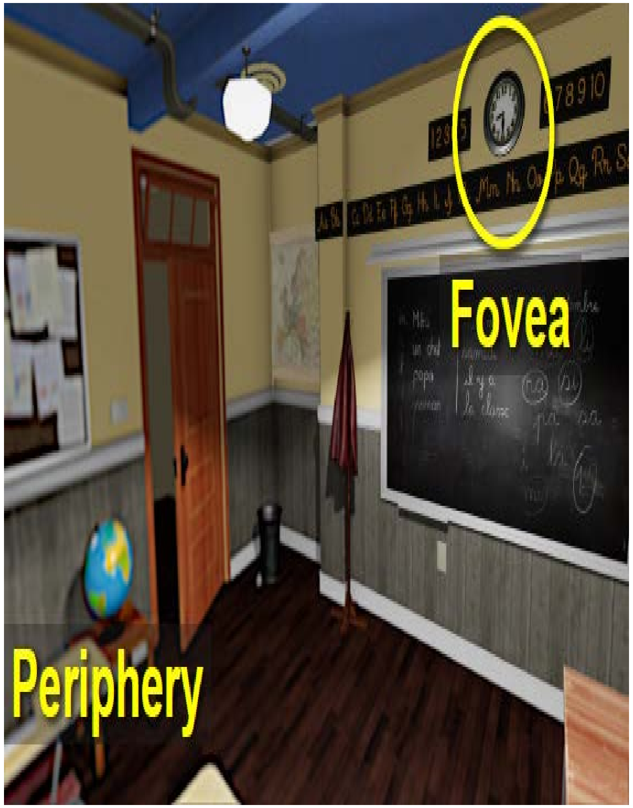
Zhiming Hu[1], Andreas Bulling[2], Sheng Li[1], Guoping Wang[1]

[1]Peking University
[2]University of Stuttgart

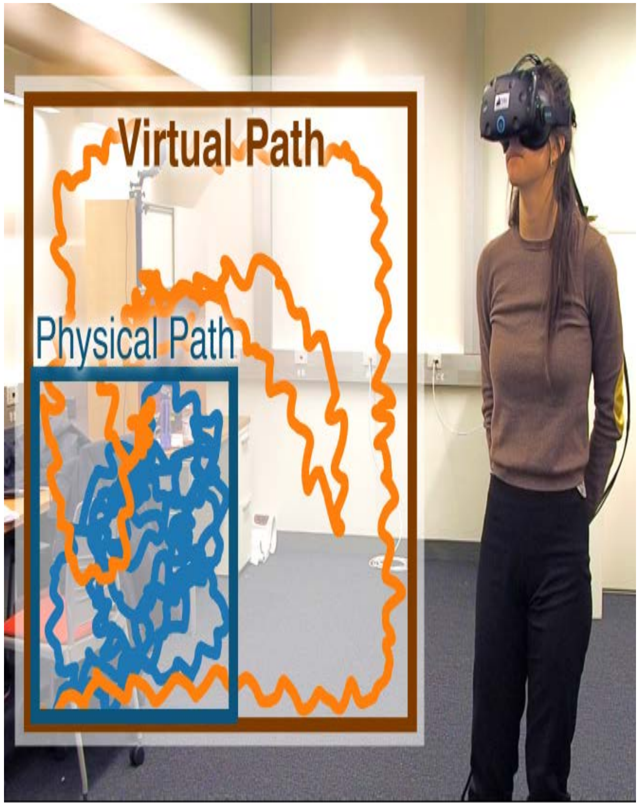
Project URL: cranehzm.github.io/EHTask

Background

Application of Human Visual Attention in VR



Gaze-Contingent Rendering
[Patney et al. 2016]



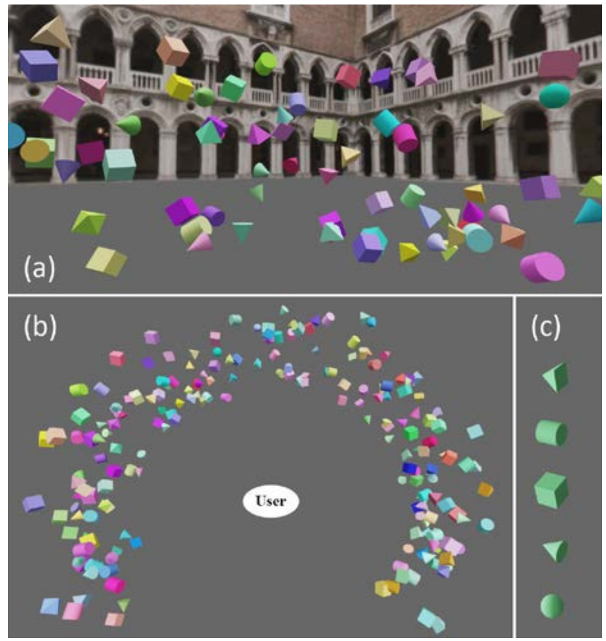
Redirected Walking
[Sun et al. 2018]

Layout Optimization
[Alghofaili et al. 2019]

Background

Application of Human Visual Attention in VR

VR Content Design
[Sitzmann et al. 2018]



Gaze Guidance
[Grogorick et al. 2017]

LOD Management
[Lee et al. 2009]

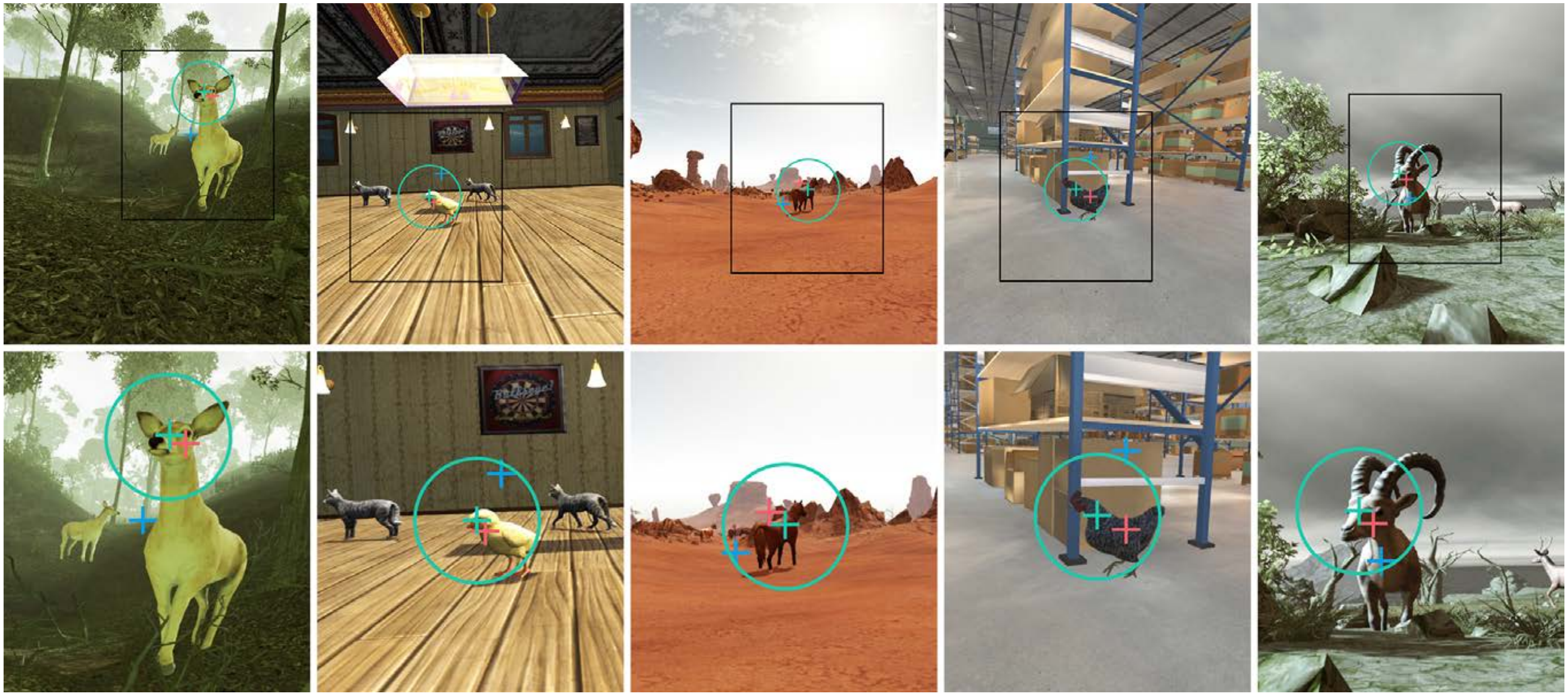
Problem Statement

- Previous works on **visual attention analysis** typically only explored **one specific VR task** and paid less attention to the **differences between different tasks**
- Existing **task recognition methods** typically focused on **2D viewing conditions** and only explored the effectiveness of **human eye movements**

Related Work

Peking University

Visual Attention Analysis in VR



Visual Attention Analysis for Free Viewing Task in VR
[Hu et al. 2020]

Visual Attention Analysis in VR

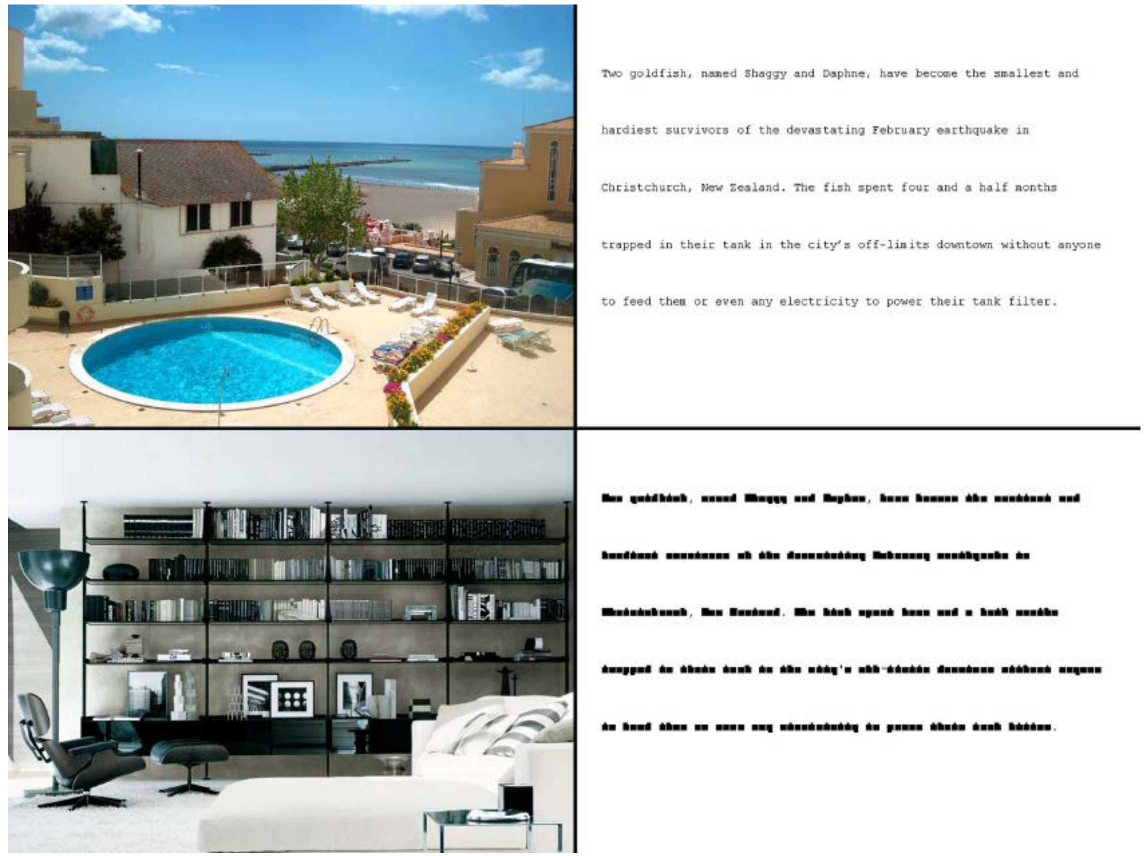
Visual Attention Analysis for Visual Search Task in VR
[Hu et al. 2021]

Related Work

PEKING UNIVERSITY

Task Recognition Methods

- Scene Memorization
- Reading
- Scene Search
- Pseudo-Reading



Task Recognition for 2D Images
[Henderson et al. 2013]

Related Work

北京大學 PEKING UNIVERSITY

Task Recognition Methods

- Explore
- Observe
- Search
- Track

Task Recognition for 2D Videos
[Hild et al. 2018]

Related Work

Compared with Prior Works on **Visual Attention Analysis:**

- Explore **four different VR tasks** using the same settings
- Analyze the **differences** between different VR tasks

Compared with Previous Works on **Task Recognition:**

- Focus on **immersive VR** instead of 2D images or videos
- Explore the effectiveness of both **human eye movements and head movements** rather than only human eye movements

Motivation

PEKING UNIVERSITY

Applications of Task Recognition Methods in VR

- **Adaptive virtual environment design**

Provide users with dynamic and adaptive experiences based on user tasks

- **Low-friction predictive interfaces**

Provide users with convenience for completing the task with less friction

- **Attention-aware intelligent systems**

Improve the usability of the system by adapting to different tasks and states of attention

Contributions

- We provide a new **dataset** that contains **human eye and head movements** under **four task conditions**
- We analyze the **patterns of human eye and head movements** and reveal **significant differences** across different tasks
- We present ***EHTask***, a novel **learning-based** method to recognize user tasks in immersive virtual reality

Data Collection

- Participants: 30 users (18 male, 12 female)
- Stimuli: 15 360-degree VR videos
- Apparatus: HTC Vive, eye tracker
- Procedure: Free viewing, Visual search, Saliency, Track
- Data: Task categories, Eye movements, Head movements

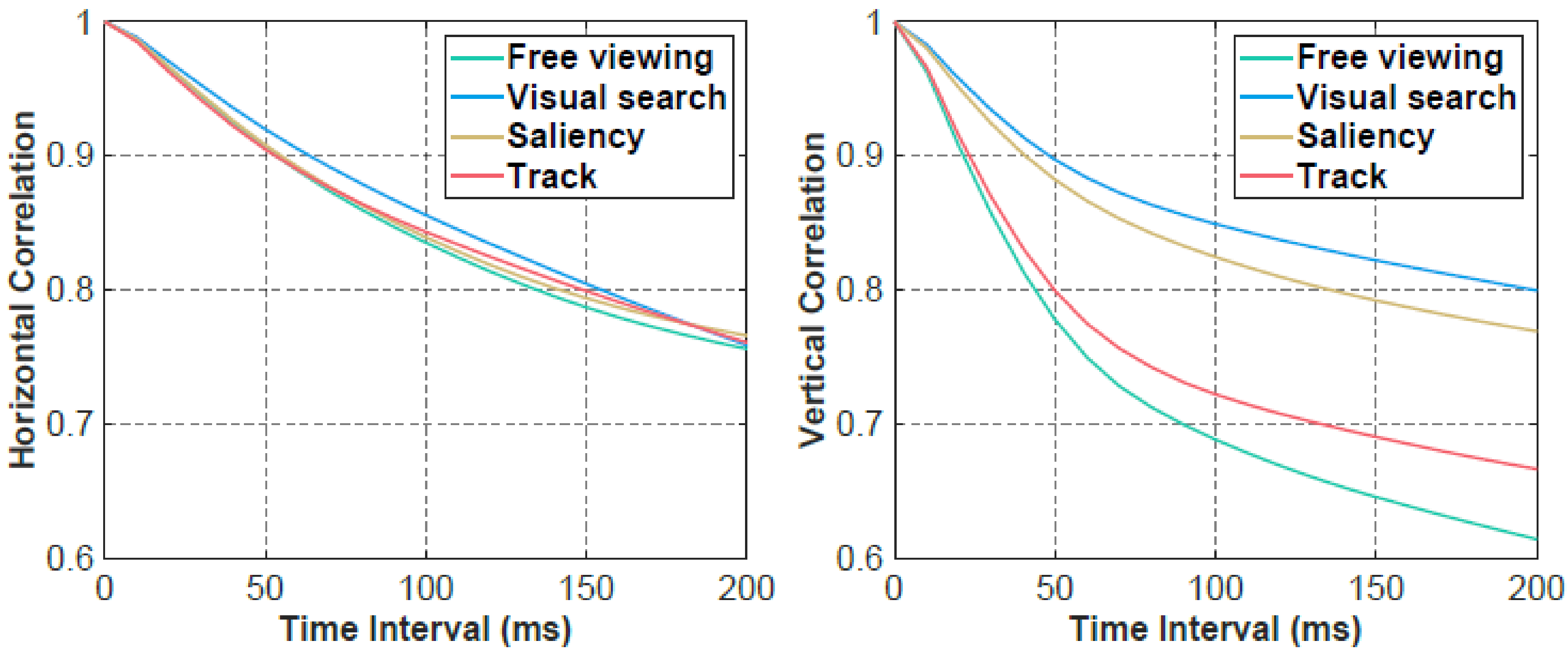
Stimuli

Statistical Characteristics of Eye Movements in the Four Tasks

		Free viewing	Visual search	Saliency	Track
Mean Fixation Duration	Mean	**263.4** ms	*339.5* ms	241.2 ms	431.7 ms
	SD	25.6 ms	49.0 ms	24.3 ms	106.7 ms
Fixation Number Per Second	Mean	**1.41**	*1.97*	1.22	1.77
	SD	0.38	0.17	0.43	0.19
Mean Saccade Duration	Mean	**633.2** ms	*269.3* ms	776.0 ms	241.1 ms
	SD	218.0 ms	69.2 ms	260.1 ms	56.2 ms
Saccade Number Per Second	Mean	**1.03**	*1.20*	0.95	**1.01**
	SD	0.17	0.18	0.19	0.24
Mean Saccade Amplitude	Mean	**6.51**°	*4.73*°	8.56°	5.40°
	SD	1.24°	1.05°	1.49°	1.58°
Fixation Distribution Dispersion	Mean	**2.21E-6**	**2.25E-6**	7.08E-6	**2.50E-6**
	SD	1.01E-6	1.18E-6	3.50E-6	1.57E-6

For each item, the difference in the fonts of two tasks indicates that there exists a significant difference between them.

Auto-Correlations of Eye Movements in the Four Tasks



The auto-correlations of the horizontal (left) and vertical (right) eye coordinates

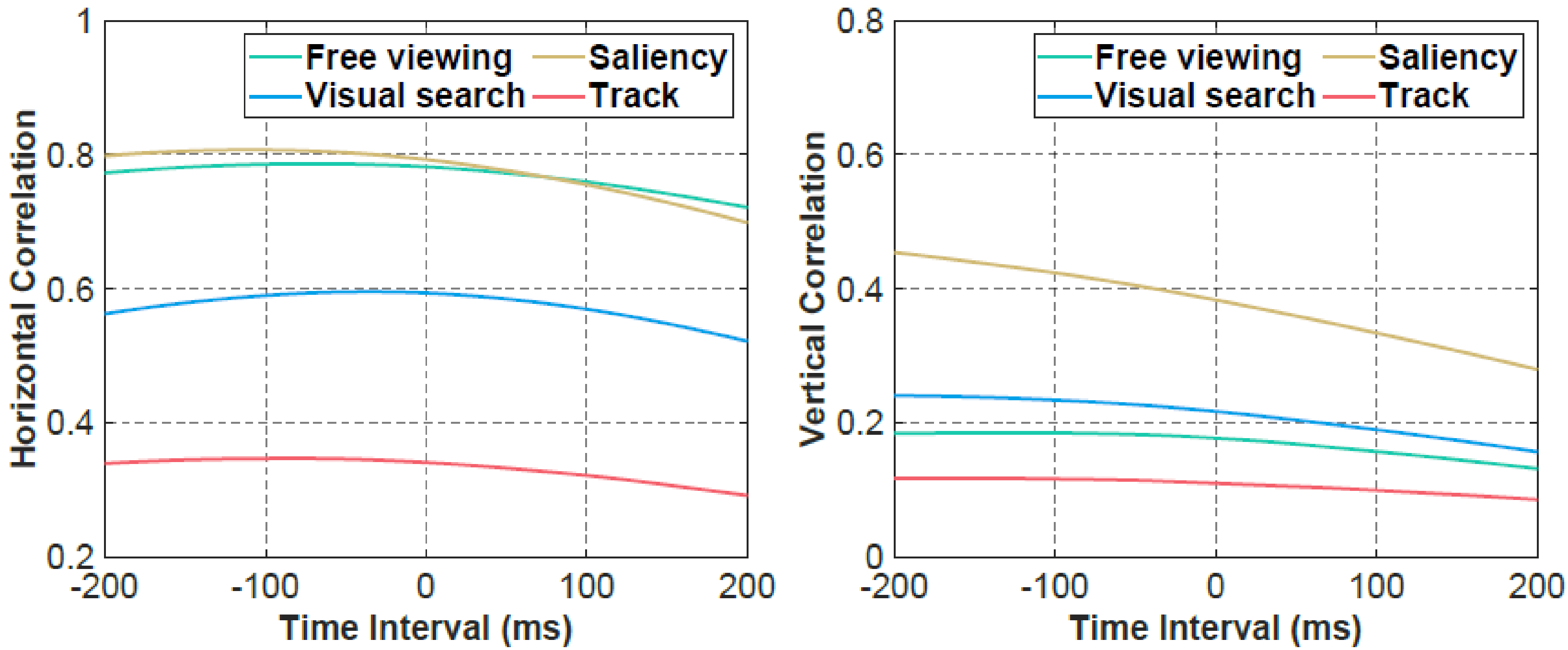
Data Analysis

Statistical Characteristics of Head Movements in the Four Tasks

		Free viewing	Visual search	Saliency	Track
Mean Absolute Horizontal Velocity	Mean	**$22.7^\circ/s$**	*$9.1^\circ/s$*	$26.8^\circ/s$	$6.4^\circ/s$
	SD	$4.3^\circ/s$	$2.3^\circ/s$	$4.4^\circ/s$	$2.4^\circ/s$
Mean Absolute Vertical Velocity	Mean	**$2.9^\circ/s$**	**$2.7^\circ/s$**	$7.5^\circ/s$	$1.9^\circ/s$
	SD	$0.6^\circ/s$	$0.5^\circ/s$	$1.4^\circ/s$	$0.4^\circ/s$
Mean Absolute Horizontal Acceleration	Mean	**$182.6^\circ/s^2$**	*$140.4^\circ/s^2$*	$203.5^\circ/s^2$	$129.8^\circ/s^2$
	SD	$29.4^\circ/s^2$	$14.1^\circ/s^2$	$23.9^\circ/s^2$	$19.4^\circ/s^2$
Mean Absolute Vertical Acceleration	Mean	**$125.0^\circ/s^2$**	*$114.2^\circ/s^2$*	$145.4^\circ/s^2$	*$109.4^\circ/s^2$*
	SD	$15.0^\circ/s^2$	$11.1^\circ/s^2$	$12.0^\circ/s^2$	$11.6^\circ/s^2$
Velocity Distribution Dispersion	Mean	**2.64E+4**	*6.95E+3*	2.39E+5	3.12E+3
	SD	2.13E+4	7.98E+3	1.27E+5	4.35E+3

For each item, the difference in the fonts of two tasks indicates that there exists a significant difference between them

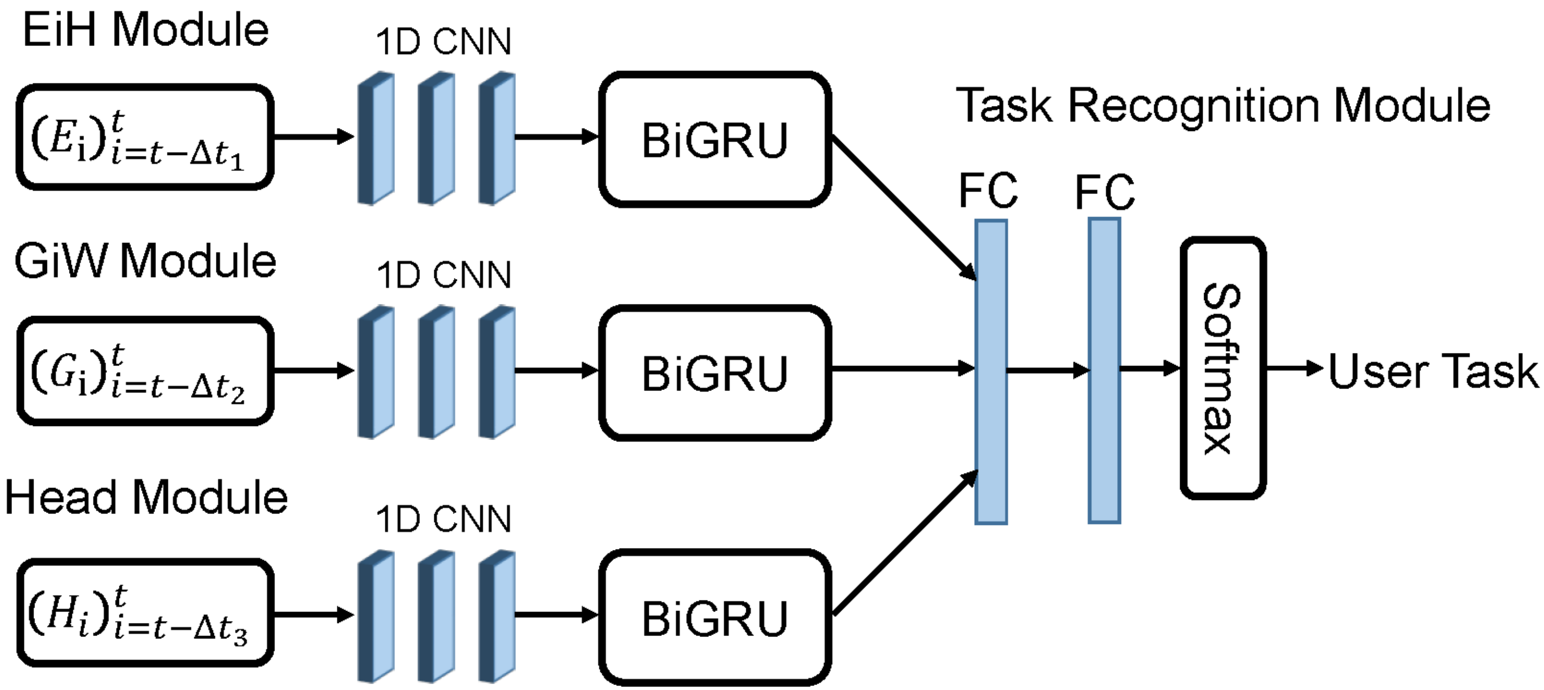
Eye-Head Coordination in the Four Tasks



The correlations between gaze positions and head velocities in the horizontal (left) and vertical (right) directions

北京大學 PEKING UNIVERSITY

EHTask Model



Architecture of EHTask model

Task Recognition Performance in **VR**

		Ours	LDA	SVM	BC	RFo	RFe
Cross-User	Window	**84.4%**	54.0%	54.3%	49.3%	<u>62.8%</u>	48.7%
	MV	**97.8%**	76.1%	75.3%	65.3%	<u>83.1%</u>	68.3%
Cross-Scene	Window	**82.1%**	53.8%	54.1%	49.0%	<u>62.6%</u>	48.3%
	MV	**96.4%**	74.2%	75.3%	64.4%	<u>83.6%</u>	72.2%

Task recognition performances on our dataset

EHTask outperforms other methods in both cross-user and cross-scene settings

Experiments and Results

Task Recognition Performance in **Real World**

	Ours	LDA	SVM	BC	RFo	RFe
Window	**61.9%**	39.0%	37.9%	36.3%	<u>44.1%</u>	36.1%
MV	**87.7%**	60.0%	46.2%	40.0%	60.0%	<u>64.6%</u>

Task recognition performances on a real-world dataset

EHTask outperforms other methods in real-world situations

PEKING UNIVERSITY

Limitations

- We only explored the **four tasks** that are most commonly used in VR applications
- We employed **non-interactive** VR videos instead of interactive 3D virtual environments as our stimuli
- We mainly focused on the differences between **different tasks** rather than the differences between **different stimuli**

PEKING UNIVERSITY

Future Work

- Overcome the **limitations**
- Explore the effectiveness of **other factors**, such as human body movements and hand movements, in recognizing user tasks
- Apply our model to **other systems**, such as real-world system, AR system, and MR system
- Recognize **other mental states** in immersive VR, such as user cognitive loads and the levels of VR cybersickness, from human eye and head movements

Thank you