PUI Perceptual User Interfaces

cbb Computational Biophysics and Biorobotics

SimTech

University of Stuttgart Germany

Towards Human-centred Artificial Intelligence

Research Talk

Zhiming Hu

January 2, 2024

Perceptual User Interfaces Group, University of Stuttgart

Computational Biophysics and Biorobotics Group, University of Stuttgart

Stuttgart Center for Simulation Science (SimTech)

Table of Contents

Education Background & Academic Positions

Education Background

- Ph.D. in Computer Software and Theory 2017.09-2022.07
  Peking University, Supervised by Prof. Guoping Wang

- B.Eng. in Optical Engineering 2013.09-2017.07
  Beijing Institute of Technology

Academic Positions

- Post-doctoral Researcher 2022.08-now
  University of Stuttgart, Led by Prof. Andreas Bulling & Prof. Syn Schmitt

Awards & Honours

As a researcher:

- Best Doctoral Student Paper Award Nominees at INTERACT 2023
- **TVCG Best Journal Award Nominees at IEEE VR 2021** (top 2%, first time for Chinese researchers)

As a student:

- **National Scholarship** (top 2%), 2021
- CSC (China Scholarship Council) Scholarship, 2020
- Chancellor's Scholarship (top 2%), 2020
- Leo KoGuan Scholarship (top 5%), 2019
- **Leader Scholarship** (top 0.2%, 7 out of over 3800 students), 2017
- **National Scholarship** (top 2%), 2016
- **National Scholarship** (top 2%), 2014

Professional Activities

Reviewing

- Journals: IMWUT, TiiS, T-MM, TVCG, IJHCI, MTAP
- Conferences: SIGGRAPH Asia, CVPR, ICCV, ECCV, CHI, UIST, IEEE VR, ISMAR

Organising committee

- Virtualisation Chair for ETRA 2024
- Associate Chair for MuC 2023
- Technical Program Committee member for iWOAR 2023

Table of Contents

Research Interests

- Human-computer interaction
- Virtual reality
- Eye tracking
- Human-centred artificial intelligence

Research goal

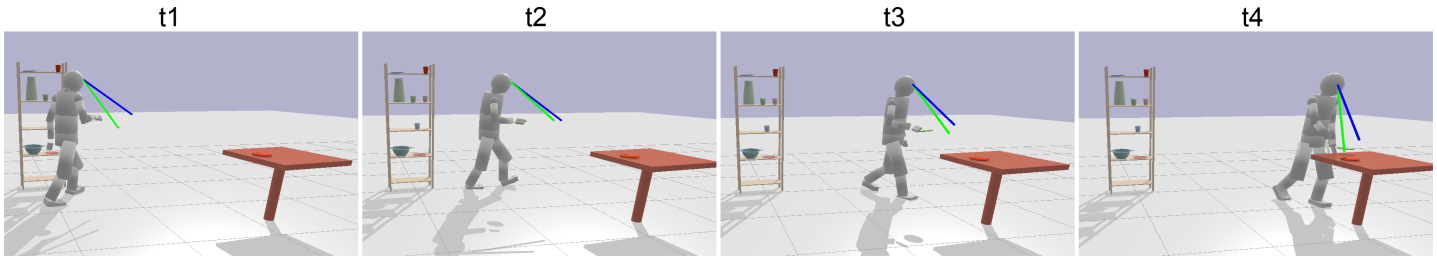
Develop human-aware intelligent user interfaces that can accurately model human daily behaviours

Towards Human-aware Intelligent User Interfaces

- Human behaviour estimation and prediction
  How to acquire human behaviour data?
- Computational human activity analysis
  What can we learn from human data?
- Human-aware intelligent system
  How to enhance the system's intelligence using human data?

Towards Human-aware Intelligent User Interfaces

- Human behaviour estimation and prediction
- Computational human activity analysis
- Human-aware intelligent system



Human daily pick and place actions

Table of Contents

Human Behaviour Estimation and Prediction

- Head pose-based gaze estimation
- Task-oriented gaze prediction
- Privacy-preserving gaze estimation
- Gaze super-resolution

Problem definition

- Input: head pose + scene content
- Output: gaze position

Static virtual environments [Hu TVCG'19]

SGaze: An Eye-head Coordination Model for Gaze Prediction

$$\tilde{x}_g = \alpha_x \cdot \tilde{v}_{hx}(t + \Delta t_x) + \beta_x \cdot a_{hx} + b_x \cdot x_S + c_x$$
$$\tilde{y}_g = \alpha_y \cdot \tilde{v}_{hy}(t + \Delta t_y) + b_y \cdot y_S + c_y$$

$\tilde{x}_g$, $\tilde{y}_g$: predicted eye gaze
$\tilde{v}_{hx}$, $\tilde{v}_{hy}$: head velocity
Δt_x, Δt_y: time interval between gaze and head
a_{hx}: horizontal head acceleration
x_S, y_S: salient positions
α_x, α_y, β_x, b_x, b_y, c_x, c_y: learned parameters

[Hu TVCG'19]

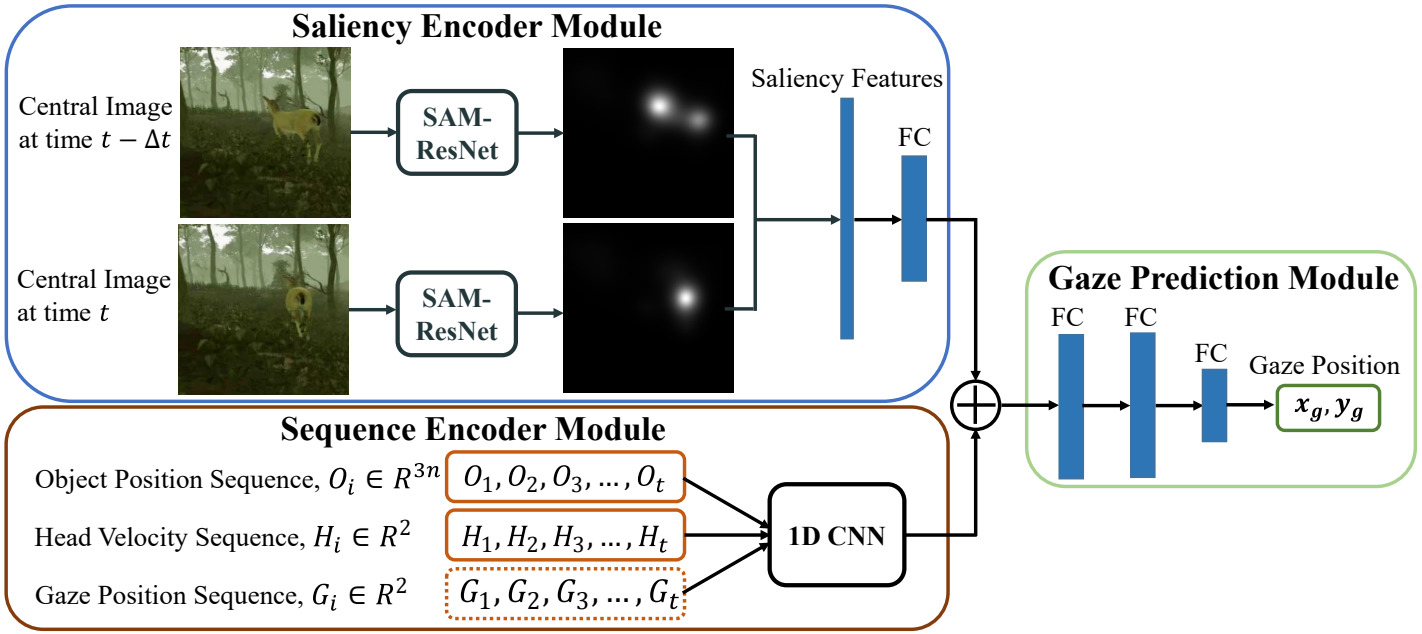
Problem definition

- Input: head pose + scene content + dynamic objects
- Output: gaze position

Dynamic virtual environments [Hu TVCG'20]

DGaze: CNN-based Gaze Prediction in Dynamic Scenes

- Gaze estimation using VR content, and head movements
- Gaze forecasting using past gaze positions



[Hu TVCG'20]

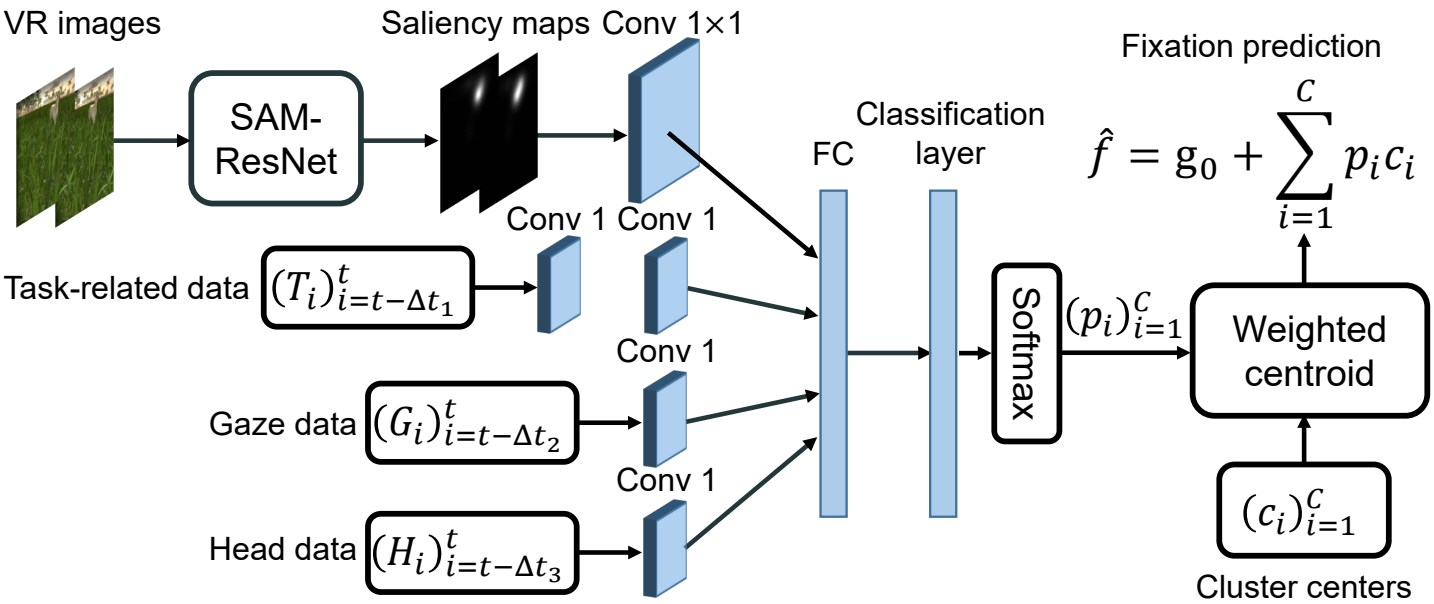
Problem definition

- Input: head pose + scene content + task-related information
- Output: future gaze fixation

Task-oriented virtual environments [Hu TVCG'21]

FixationNet: Gaze Forecasting in Task-oriented Environments

- Extract features from VR content, past gaze and head data
- Forecast fixation using prior knowledge of gaze distribution



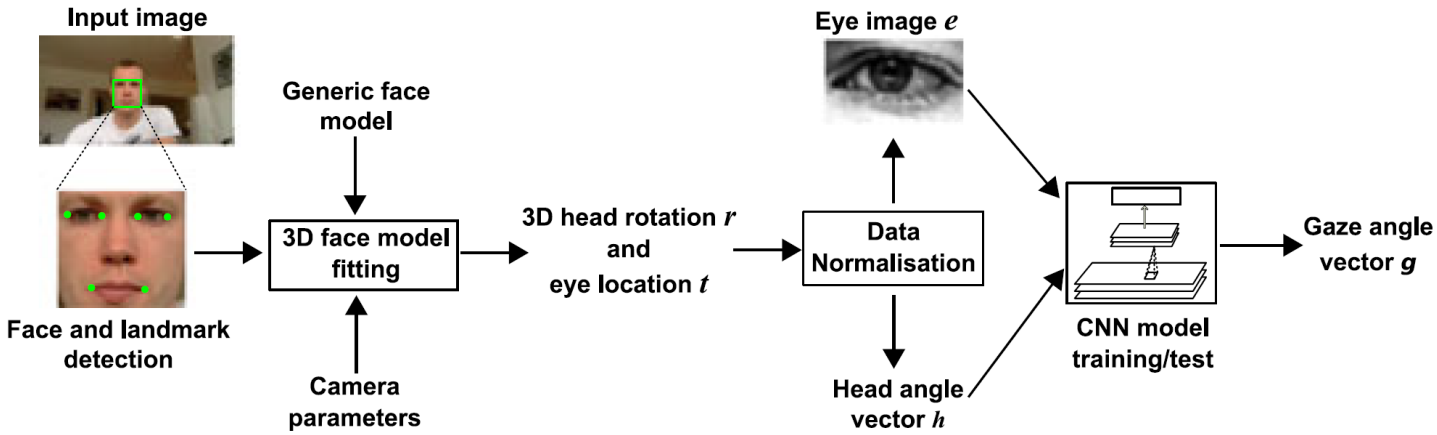
[Hu TVCG'21 **Best Journal Nominees Award**]

Background

- Gaze estimation from face or eye images
- Face or eye images are privacy-sensitive

Problem definition

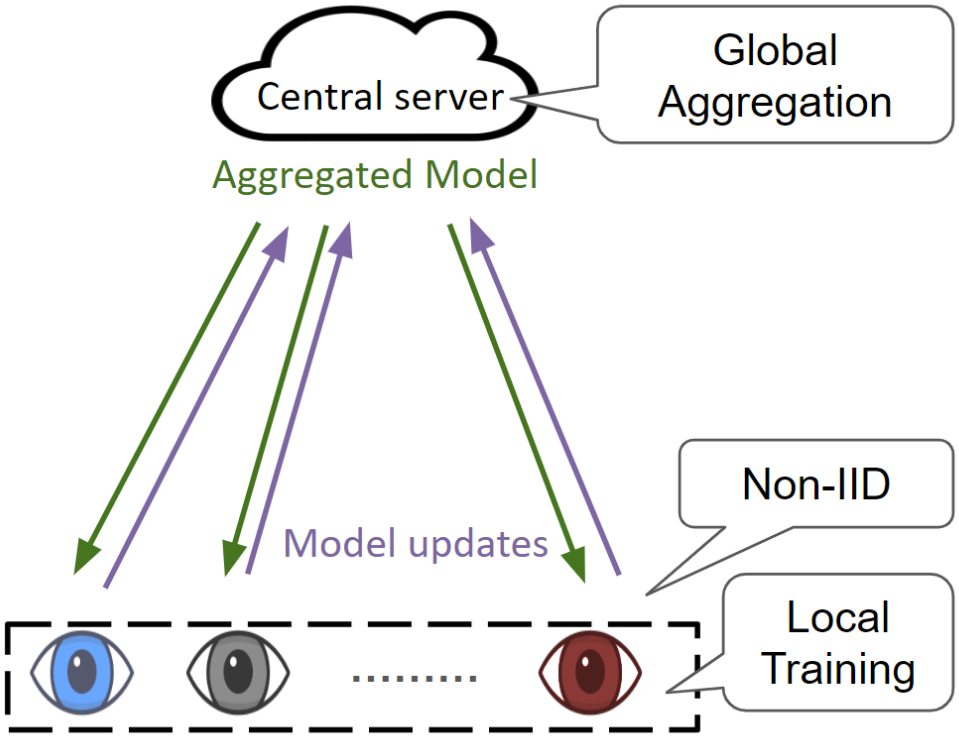
- Enhance the privacy of appearance-based gaze estimators



Appearance-based gaze estimation [Zhang PAMI'17]

Privacy-preserving Gaze Estimation using Federated Learning

- Local training to preserve privacy
- Global aggregation to ensure accuracy



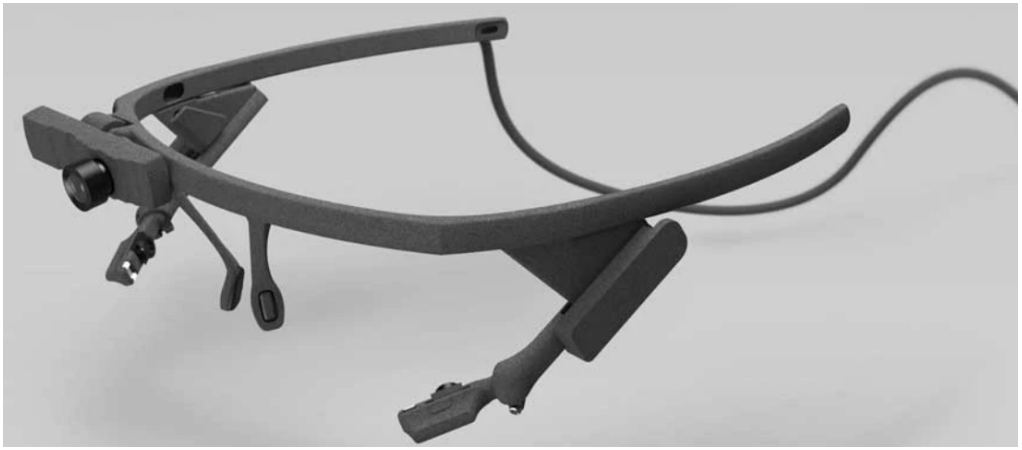
Gaze Super-resolution

Background

- Mobile eye trackers usually suffer from low-resolution
- High-resolution gaze data is significant for many applications

Problem definition

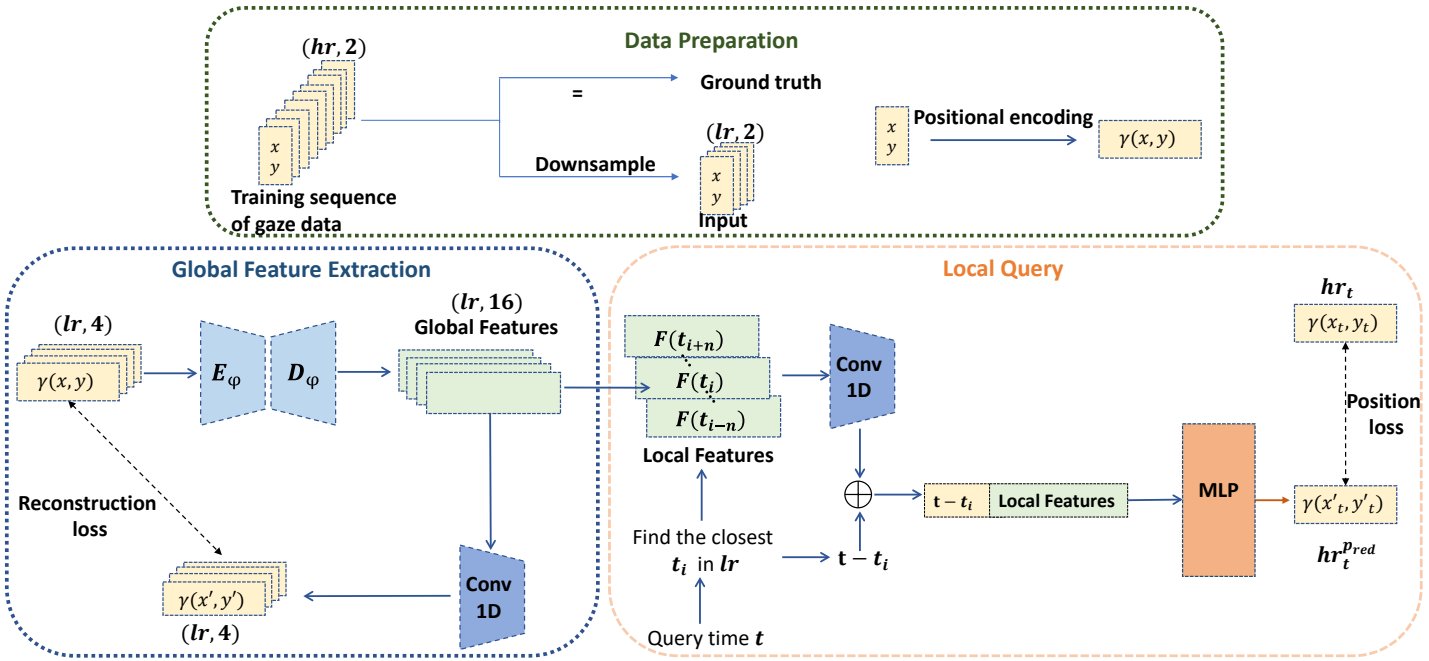
- Input: low(er)-resolution gaze data
- Output: high(er)-resolution gaze data



Mobile eye tracker [https://pupil-labs.com/products/core]

SUPREYES: SUPer Resolution for EYES

- Implicit neural representation learning
- Global feature extraction and local query



[Jiao UIST'23]

Summary

- Head pose-based gaze estimation
- Task-oriented gaze prediction
- Privacy-preserving gaze estimation
- Gaze super-resolution

Future work

- Human pose estimation
- Human motion prediction
- Hand pose estimation

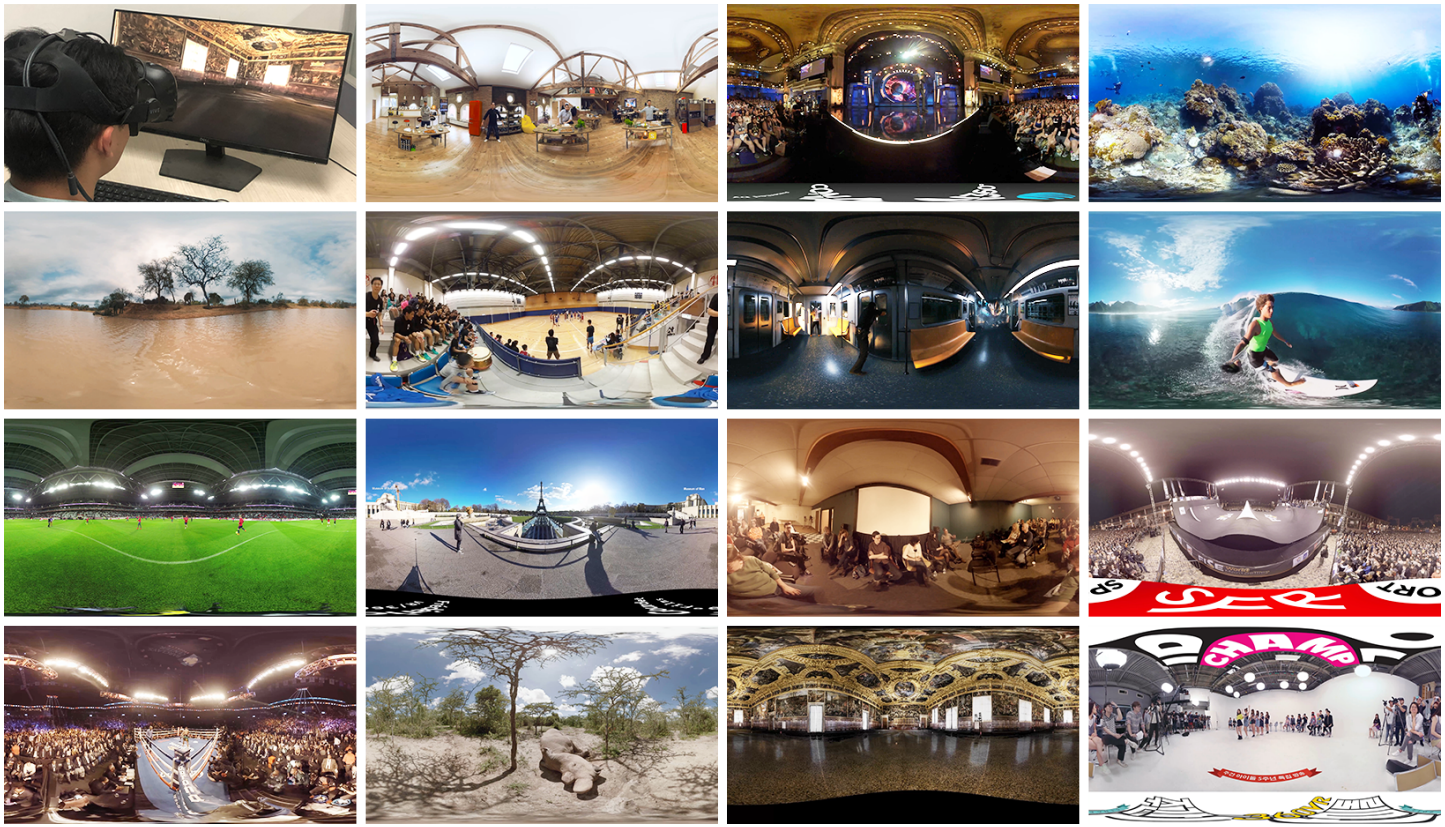
Table of Contents

Computational Human Activity Analysis

- Eye and head movement analysis
- Mouse and keyboard behaviour analysis

Problem definition

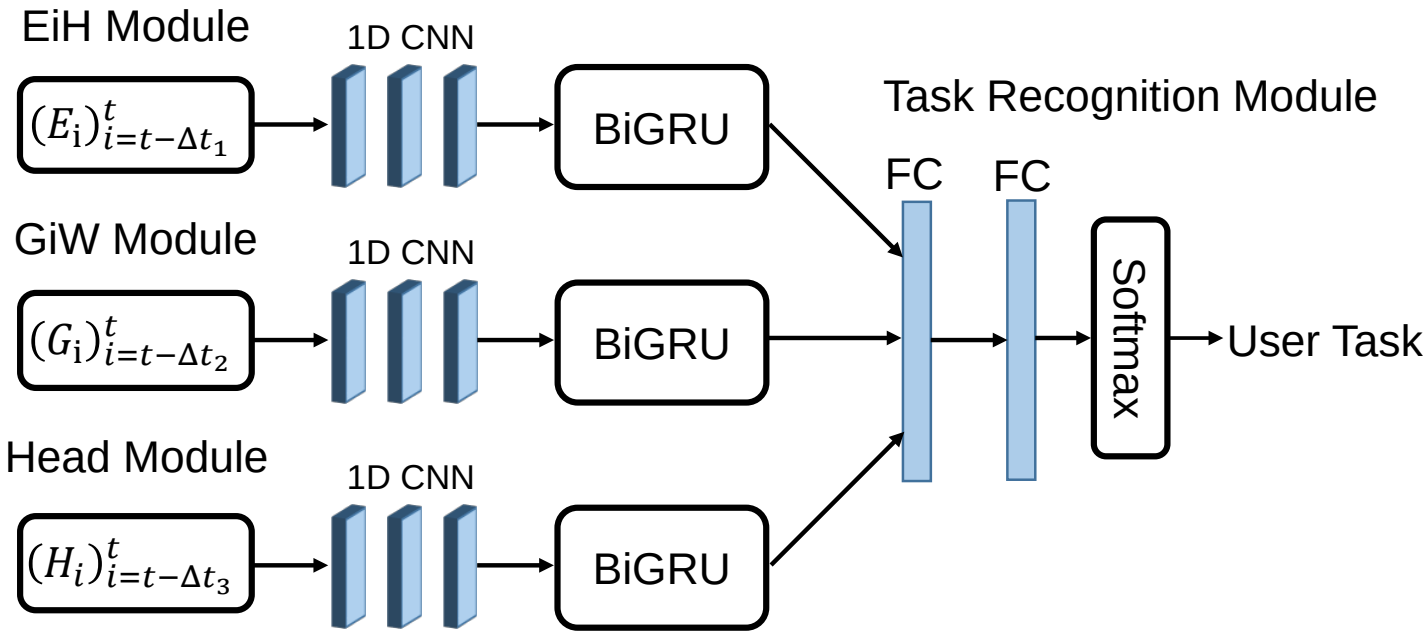
- Analyse eye and head movements under different tasks
- Recognise user tasks from eye and head features



360-degree VR videos [Hu TVCG'22]

EHTask: Task Recognition from Eye and Head Movements

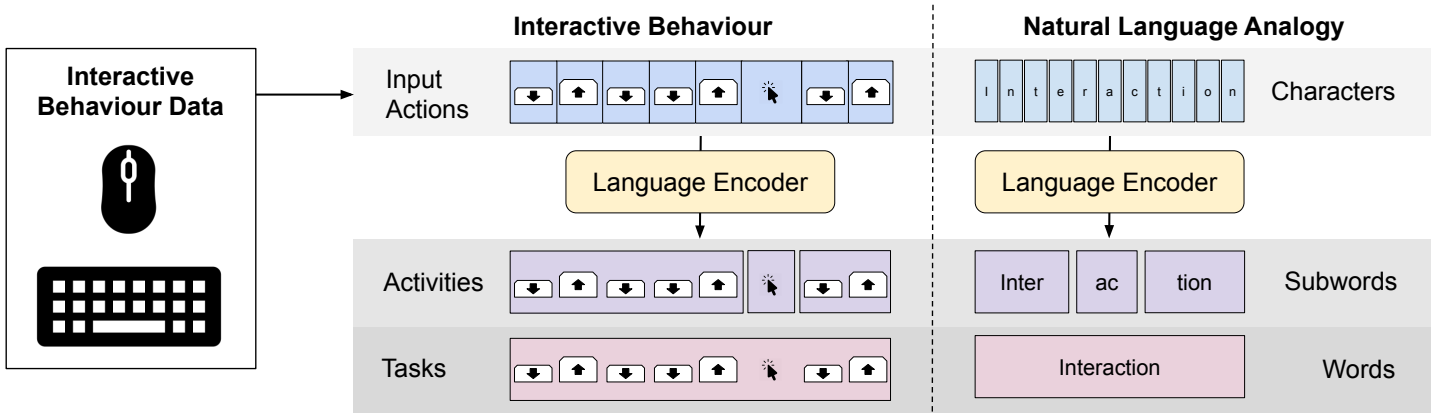
- Extract features from eye and head movements
- Fuse eye and head features to recognise user tasks



[Hu TVCG'22]

Background

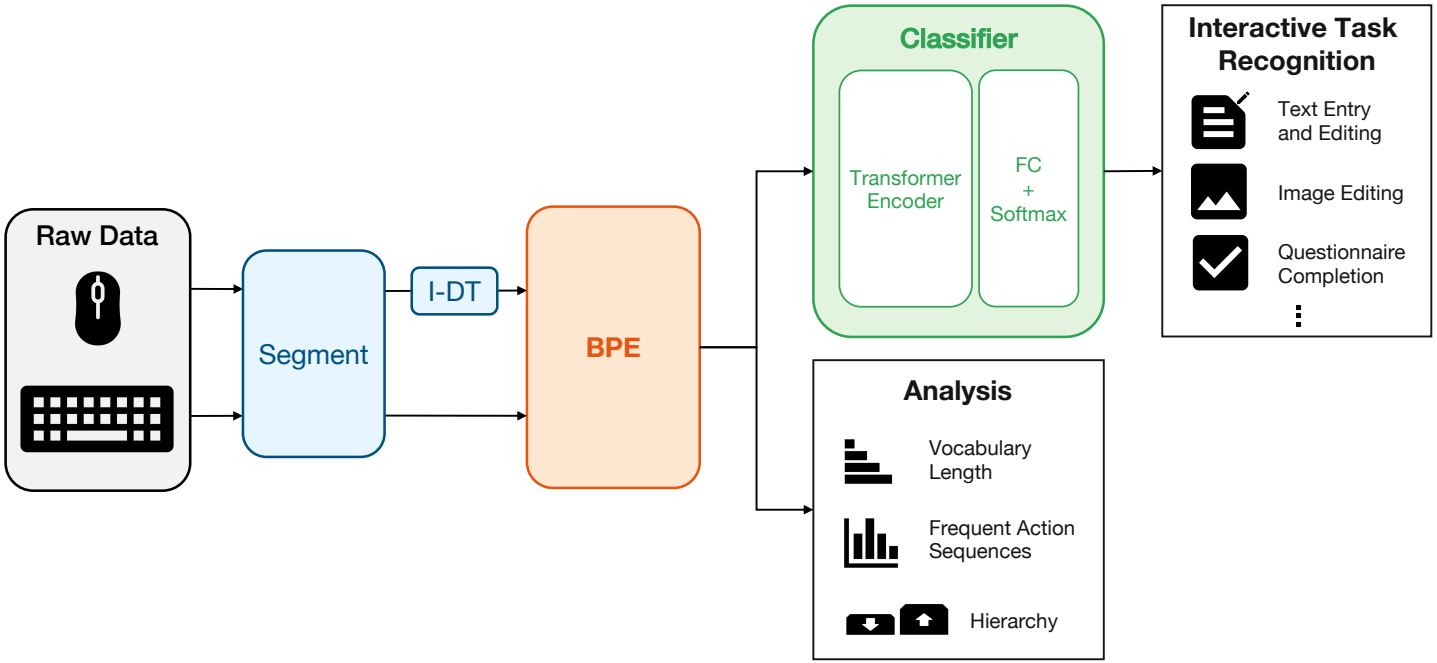
- Interactive behaviour is similar to natural language
- Can NLP methods be used to model interactive behaviour?



[Zhang INTERACT'23 **Best Student Paper Nominees**]

Modelling Interactive Behaviour using NLP Methods

- Byte pair encoding (BPE) to encode mouse and keyboard behaviour
- Transformer to recognise user tasks



[Zhang INTERACT'23 **Best Student Paper Nominees**]

Summary

- Eye and head movement analysis
- Mouse and keyboard behaviour analysis

Future work

- Human motion analysis
- Human interaction intention analysis

Table of Contents

Human-aware Intelligent System

- Gaze-contingent rendering system
- Head-assisted locomotion system

Background

- Virtual reality system requires high refresh rate to ensure user experience
- High refresh rate is computationally expensive

VR rendering [Hu TVCG'20]

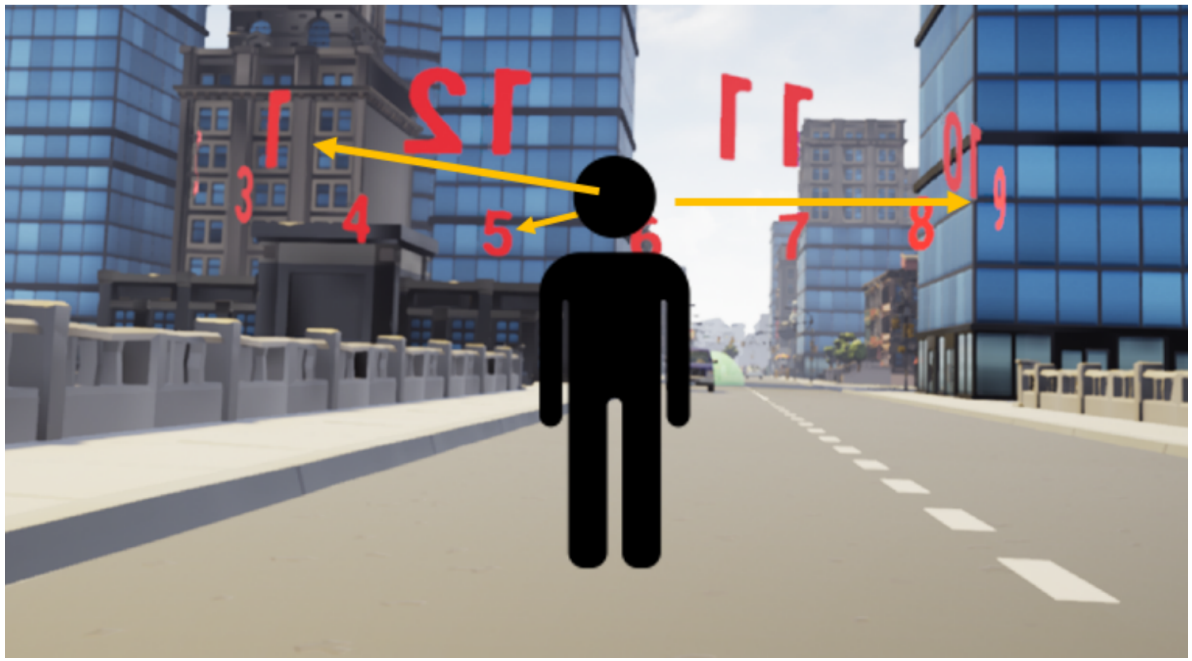
Gaze Estimation for Gaze-contingent Rendering

- Estimate eye gaze in virtual environments
- Apply estimated eye gaze to gaze-contingent rendering

Gaze-contingent rendering [Hu TVCG'20]

Background

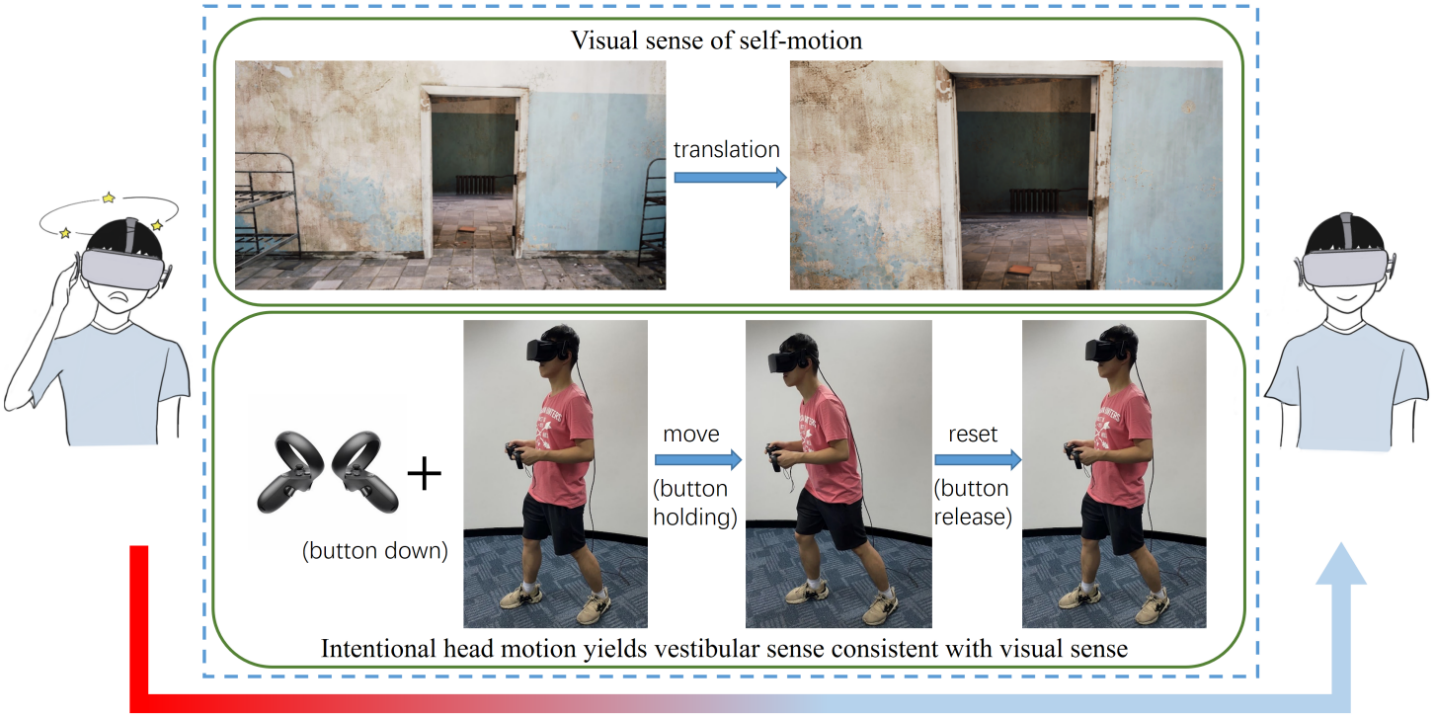
- Locomotion is important for exploring virtual environments
- Cybersickness happens during locomotion



Locomotion in VR [Lin TVCG'22]

Intentional Head Motion-assisted Locomotion

- Cybersickness is correlated with self-motion velocity
- Intentional head motion can reduce cybersickness



[Lin TVCG'22]

Summary

- Gaze-contingent rendering system
- Head-assisted locomotion system

Future work

- Intention-aware adaptive system
- Low-friction predictive interface

Table of Contents

Conclusion

Towards Human-aware Intelligent User Interfaces

- Human behaviour estimation and prediction
- Computational human activity analysis
- Human-aware intelligent system

Q & A

Any question?

Acknowledgement

Thank you!

References i

Elfares NeurIPS GMML'22. Federated learning for appearance-based gaze estimation in the wild. In *Proceedings of the NeurIPS Workshop Gaze Meets ML (GMML)*, pages 1–17, 2022. doi: 10.48550/arXiv.2211.07330.

Hu TVCG'19. Sgaze: a data-driven eye-head coordination model for realtime gaze prediction. *IEEE Transactions on Visualization and Computer Graphics*, 25(5):2002–2010, 2019.

Hu TVCG'20. Dgaze: Cnn-based gaze prediction in dynamic scenes. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):1902–1911, 2020.

Hu TVCG'21. Fixationnet: forecasting eye fixations in task-oriented virtual environments. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2681–2690, 2021.

Hu TVCG'22. Ehtask: recognizing user tasks from eye and head movements in immersive virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

Jiao UIST'23. Supreyes: Super resolution for eyes using implicit neural representation learning. In *Proc. ACM Symposium on User Interface Software and Technology*, pages 1–13, 2023. doi: 10.1145/3586183.3606780.

Lin TVCG'22. Intentional head-motion assisted locomotion for reducing cybersickness. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

Zhang INTERACT'23. Exploring natural language processing methods for interactive behaviour modelling. In *Proc. IFIP TC13 Conference on Human-Computer Interaction (INTERACT)*, pages 1–22. Springer, 2023.

Zhang PAMI'17. Mpiigaze: real-world dataset and deep appearance-based gaze estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1):162–175, 2017.