SGaze: A Data-Driven Eye-Head Coordination Model for Realtime Gaze Prediction

Zhiming Hu[1], Congyi Zhang[1], Sheng Li[1]
Guoping Wang[1], Dinesh Manocha[2]

[1]Peking University
[2]University of Maryland

Project Website: https://cranehzm.github.io/SGaze

Motivation

Eye Tracking in Virtual Reality

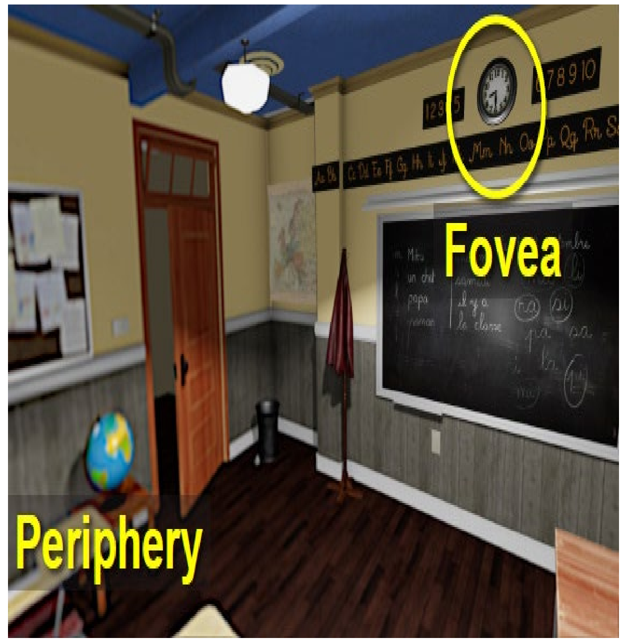
Eye Tracking[1]

[1] https://www.7invensun.com/

Motivation

Eye Tracking in Virtual Reality

VR content design
[Sitzmann et al 2018]



Gaze-contingent rendering
[Patney et al 2016]

Gaze based interaction
[Pfeiffer et al 2008]

Motivation

Solution to Eye Tracking in VR

Hardware Solution

Eye Tracker[1]

- Accurate ✓
- Expensive ✗
- Lack ease of use ✗

Software Solution?

[1] https://www.7invensun.com/

Related Work

Salient Object Detection

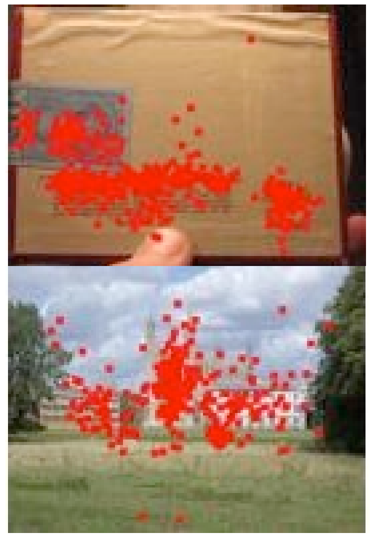
Top: original images; Bottom: salient objects[1]

[1] https://mmcheng.net/msra10k/

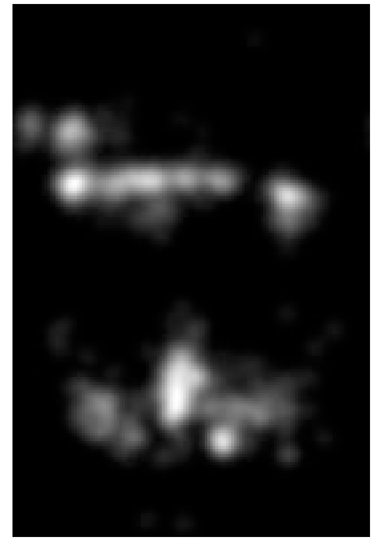
Related Work

Saliency Prediction

Original Image[1]



Eye Fixation[1]



Saliency Map[1]

Our goal: predict realtime gaze position!

[1] http://saliency.mit.edu/results_mit300.html

Contributions

- Propose a novel eye-head coordination model (SGaze)
- Propose a novel gaze prediction method based on our model
- Build a dataset for gaze prediction and provide a thorough analysis of our dataset
- Apply our model to gaze-contingent rendering

Talk Outline

- Data collection
- Data analysis
- Eye-head coordination model
- Results
- Limitations, and Future Work

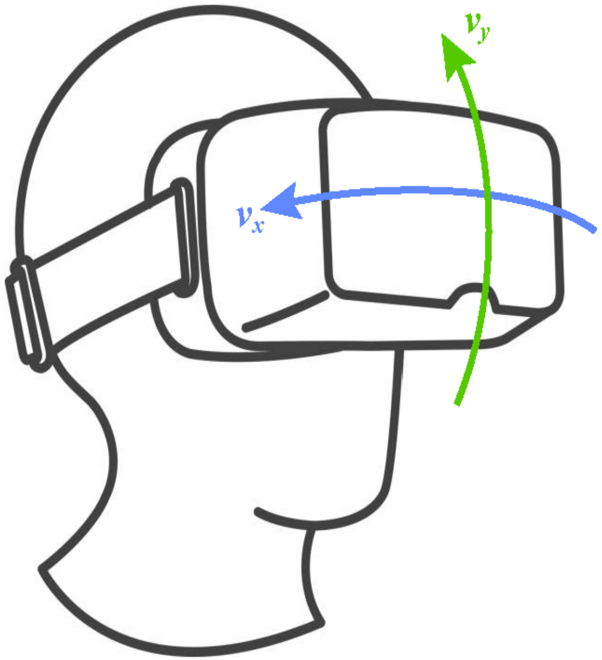
Data Collection

- Participants: 60 users (35 male, 25 female, ages 18-36)
- Stimuli: 7 scenes, static and soundless
- System: HTC Vive + eye tracker
- Procedure: free exploration, no task
- Data: realtime scenes + gaze positions + head poses

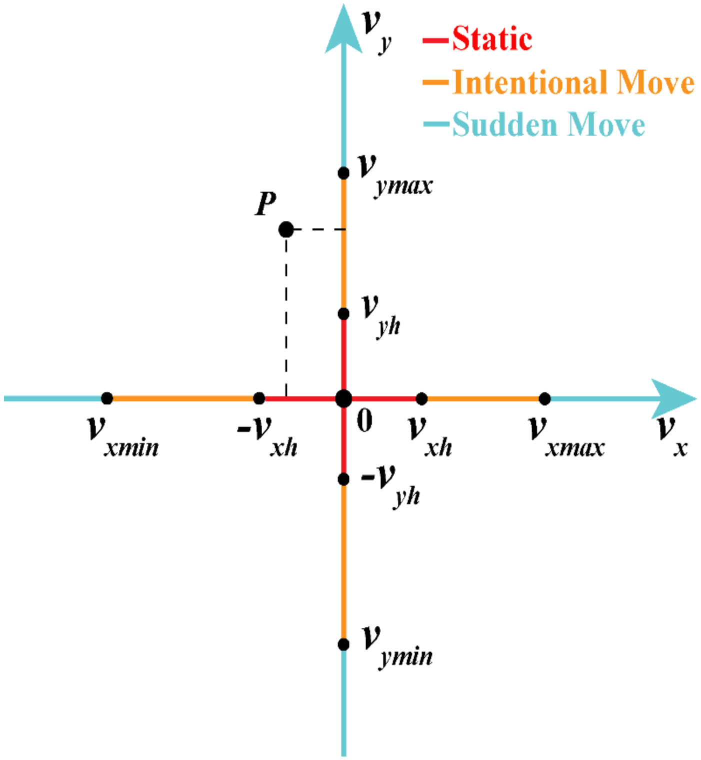
Stimuli

Data Collection

Data Analysis: Head Movement



Head velocity coordinate system



Three regions of head velocity

Data Analysis: Head Movement

	Static	Intentional	Sudden
Horizontal	5.55%	91.45%	3.00%
Vertical	4.54%	90.69%	4.77%

Distribution of data in different regions

Most of the data lies in Intentional Move region.

Data Analysis: Eye-Head Linear Correlation

Pearson's correlation coefficient (PCC)

	Static	Intentional	Sudden	Whole
PCC(v_x)	0.1345	0.5883	0.1511	0.5641
PCC(v_y)	0.1484	0.4969	-0.0906	0.4132

The PCCs between gaze position and head velocity in different regions

Head rotation velocity has a strong linear correlation with gaze position in a certain range.

Turn left/right head → Look left/right

Turn up/down head → Look up/down

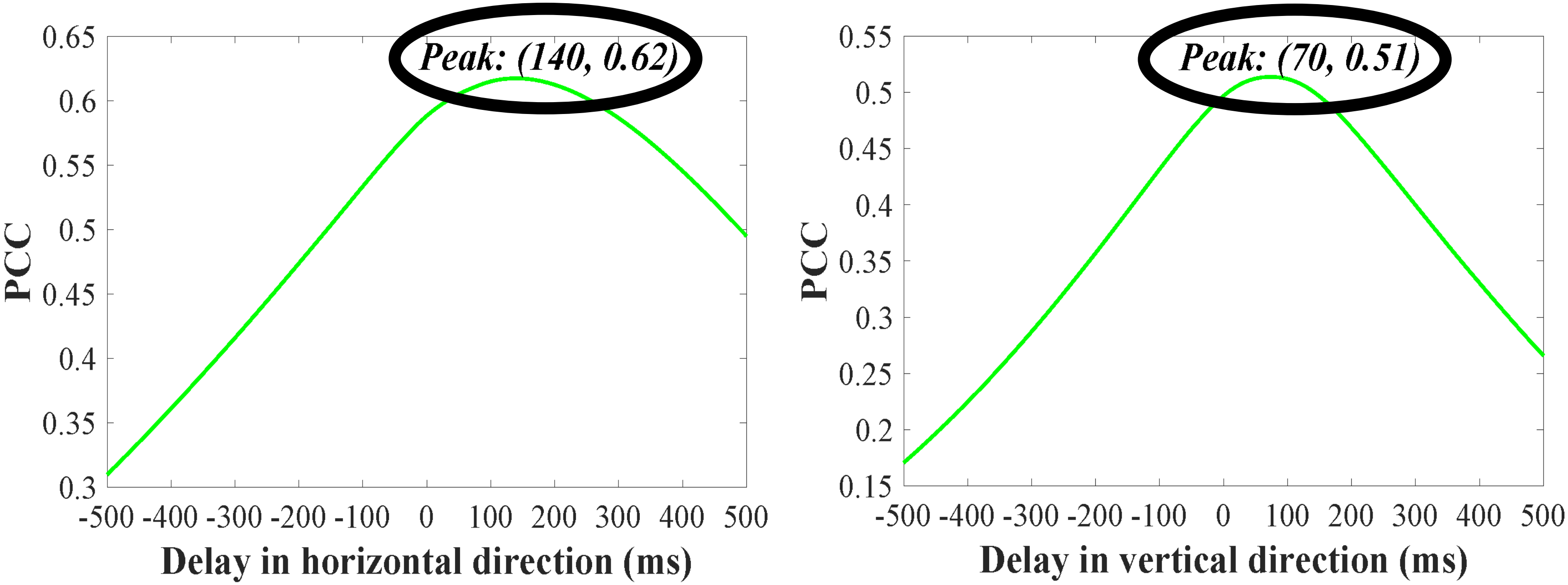
Data Analysis: Eye-Head Linear Correlation

PCC(v_x)	0.5641	PCC(a_x)	0.1134
PCC(v_y)	0.4132	PCC(a_y)	0.0132

Left: the PCCs between gaze position and head velocity
Right: the PCCs between gaze position and head acceleration

Eye-head linear correlation is stronger in the horizontal direction than in the vertical direction.

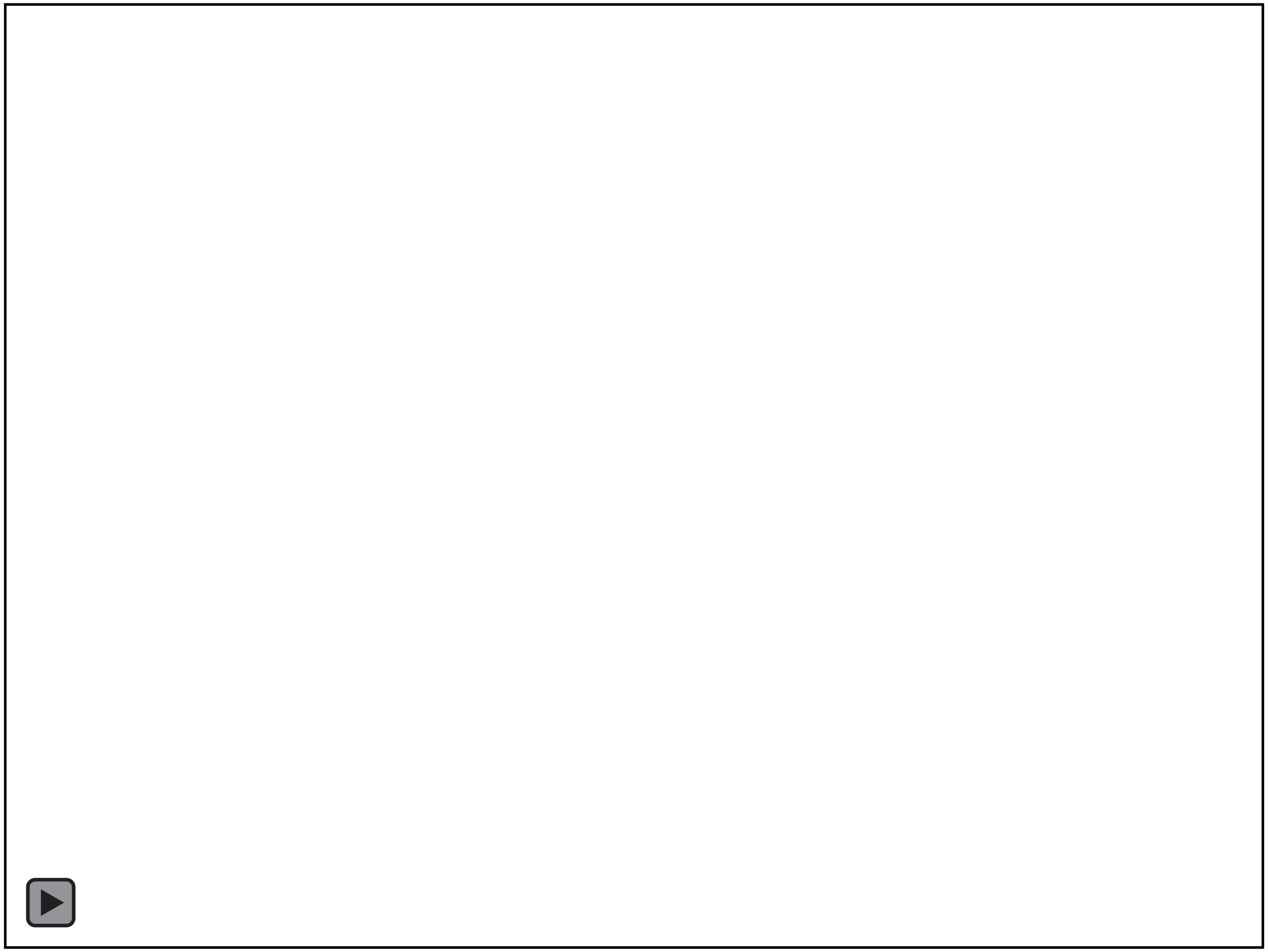
Data Analysis: Eye-Head Latency



The latencies between eye movements and head movements in horizontal (left) and vertical (right) directions

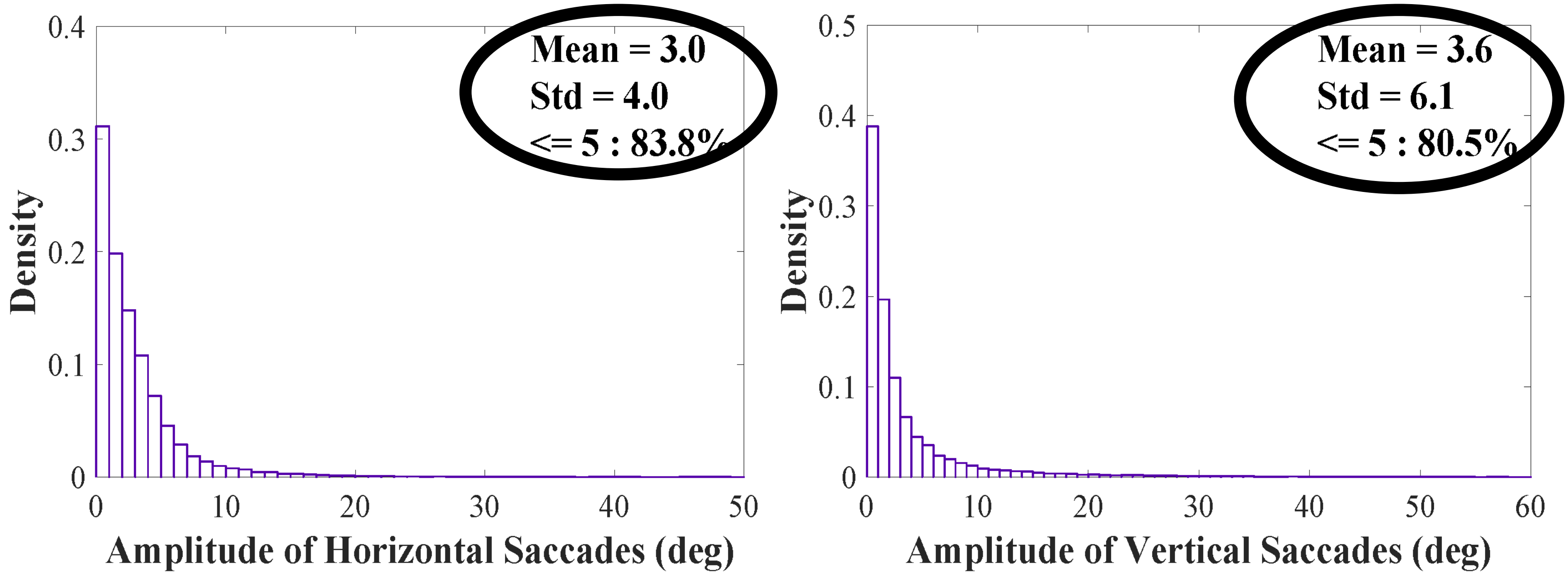
Eye movements usually happen before head movements.

Data Analysis: Saccade Analysis



Data Analysis: Saccade Analysis

Saccade: fast eye movement



Amplitudes of horizontal (left) and vertical (right) saccades

Long saccades seldom occur in
free exploration condition.

Eye-Head Coordination Model (SGaze)

Gaze = Head + Content + Task

$$x_g(t) = \alpha_x \cdot v_{hx}(t + \Delta t_{x1}) + \beta_x \cdot a_{hx}(t) + F_x(t + \Delta t_{x2}) + G_x(t) + H_x(t)$$
$$y_g(t) = \alpha_y \cdot v_{hy}(t + \Delta t_{y1}) + F_y(t + \Delta t_{y2}) + G_y(t) + H_y(t)$$

x_g, y_g: gaze position

v_{hx}, v_{hy}, a_{hx} : head velocity and acceleration

F_x, F_y: content

G_x, G_y: task

H_x, H_y: other factors

α_x, α_y, β_x: the linear influence of velocity and acceleration

Δt_{x1}, Δt_{y1}: eye-head latencies

Eye-Head Linear Correlation

Eye-Head Latency

Results

Baselines: center, mean, salient position

Evaluation Metrics: angular distance, precision and recall rates

Performance Evaluation

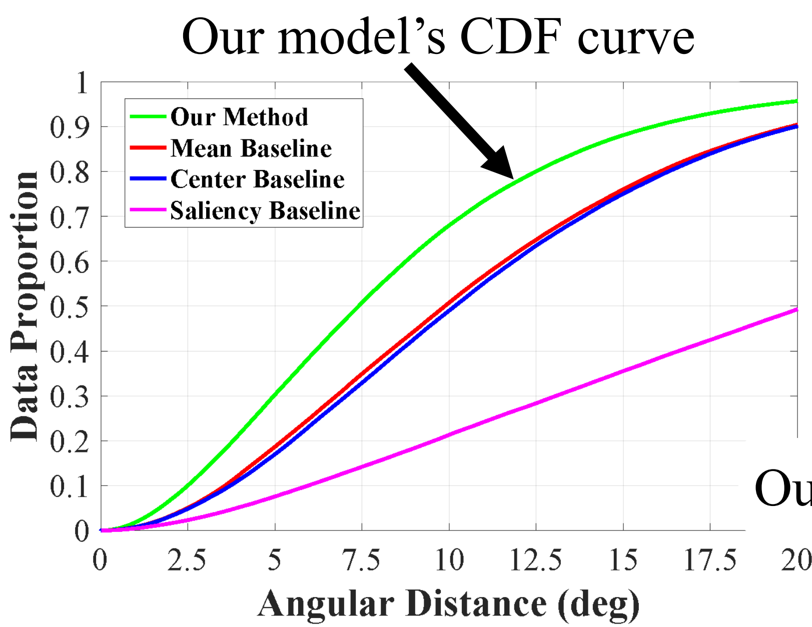
	Ours	Mean	Center	Saliency
Mean	8.52°	10.93°	11.16°	21.23°
Std	5.66°	6.43°	6.44°	12.10°

Comparison of angular distance between our model and the baselines.

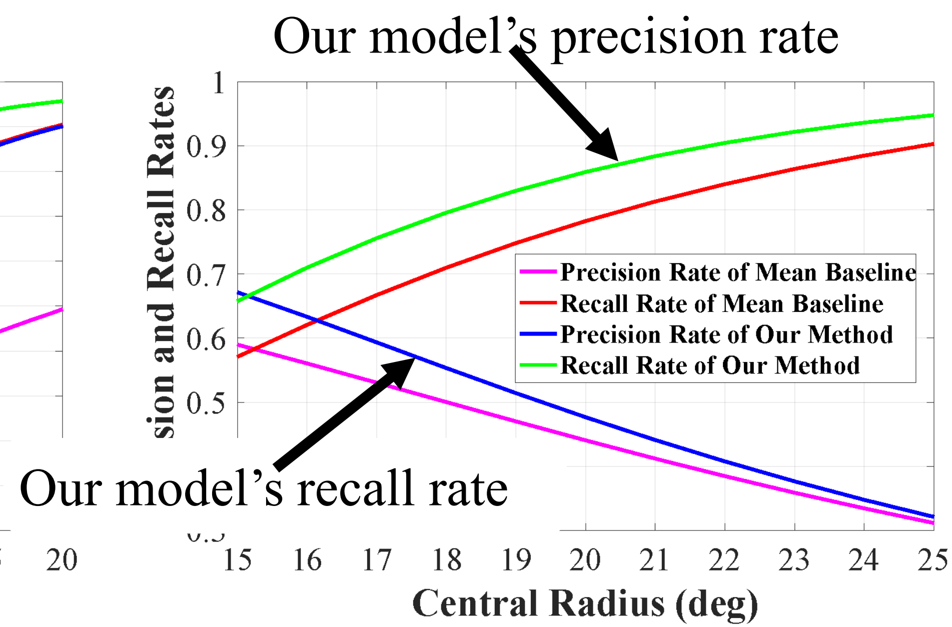
Our model performs best in terms of both mean and standard deviation.

Results

Performance Evaluation



Cumulative distribution function (CDF) of the angular distance.



Precision and recall rates at different central radii.

Results

Ablation Study

	Ours	w/o Saliency	w/o Latency	w/o Velocity	Mean
Mean	8.42°	8.48°	8.49°	10.85°	10.96°
Std	5.63°	5.66°	5.65°	6.36°	6.42°

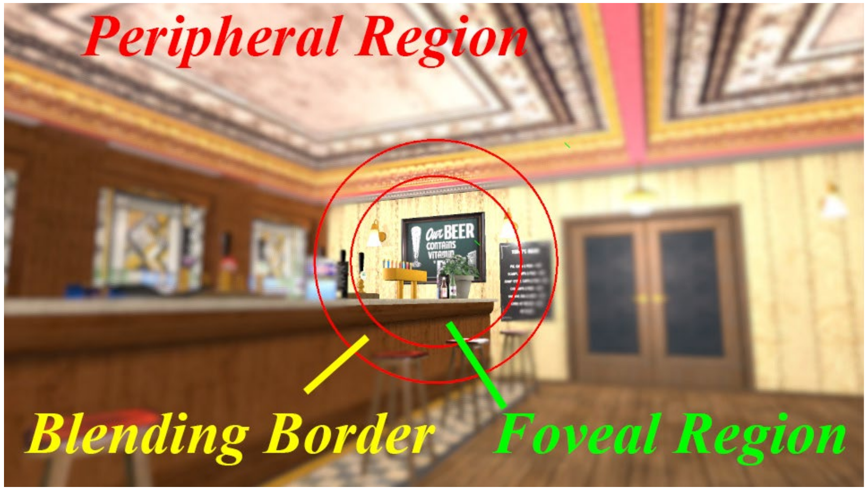
Angular distances of the ablated models.

Each component in our model contributes to gaze prediction.

Results

Gaze-Contingent Rendering

Normal mode



Gaze-contingent rendering

User Study

Ours *vs* Baseline

t-test, $p < 0.01$

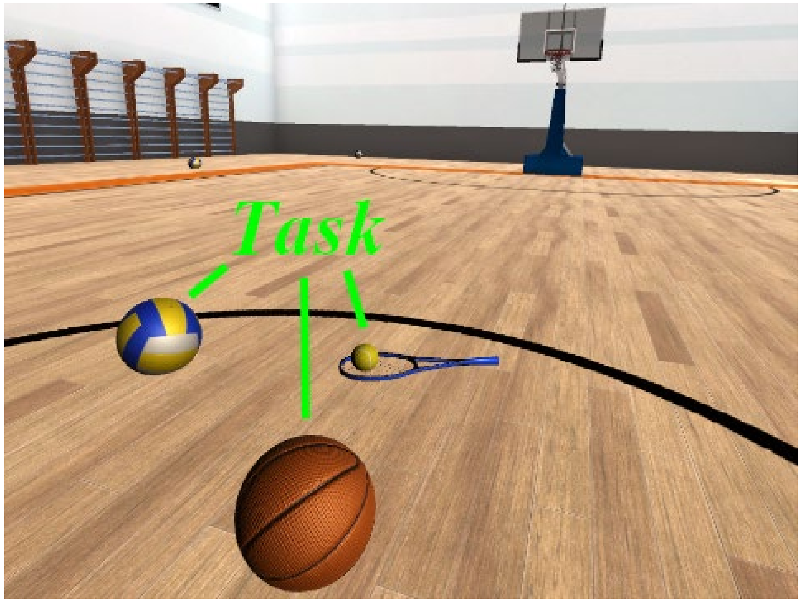
Our model is significantly better than the baseline.

Gaze-Contingent Rendering

Performance on Simple Task

Simple task

Count trees



Look for balls

Performance on Simple Task

Result

	Ours	Mean	Center	Saliency
Mean	8.99°	10.48°	10.69°	18.49°
Std	5.76°	6.00°	6.03°	13.11°

Comparison of angular distance between our model and the baselines for the simple tasks.

Our model still outperforms the baselines when there exists a simple task.

Limitations and Future Work

Limitations

- Free exploration condition (no-task situation)
- Soundless situation
- Static scenes

Future Work

- Task-oriented situation
- Sound
- Dynamic scenes
- Deep Learning

Take-Home Message

- Head pose data can facilitate gaze prediction.
- Head rotation velocity has a strong linear correlation with gaze position in a certain range.
- Eye movements usually happen before head movements.
- Gaze-contingent rendering can be achieved using our model.

Homepage: https://cranehzm.github.io/
Project Website: https://cranehzm.github.io/SGaze