

Investigating the Coordination of Human Eye Gaze and Body Movements in Extended Reality

GAMES Webinar 2025

Zhiming Hu

University of Stuttgart, Germany

zhiminghu.net ↗



Table of Contents

Research Background

Related Work

Coordination of Eye Gaze and Full-Body Poses

Coordination of Eye Gaze and Hand Gestures

Discussion

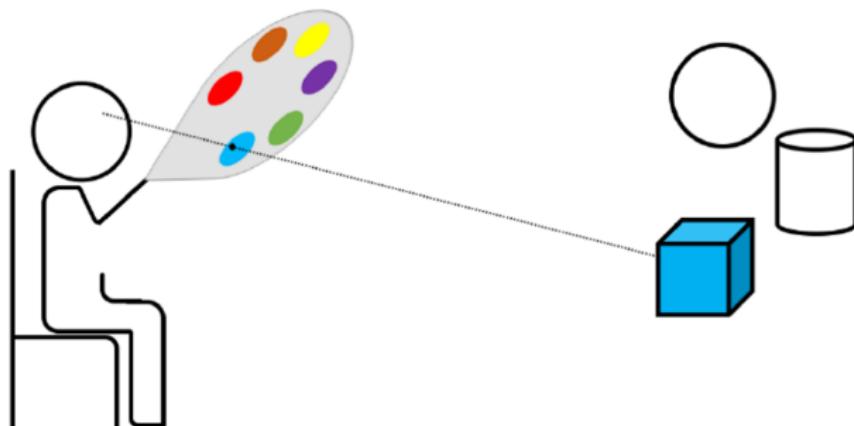
Conclusion

Applications of human eye gaze in XR



Gaze-contingent rendering [Hu TVCG'20]

Applications of human eye gaze in XR



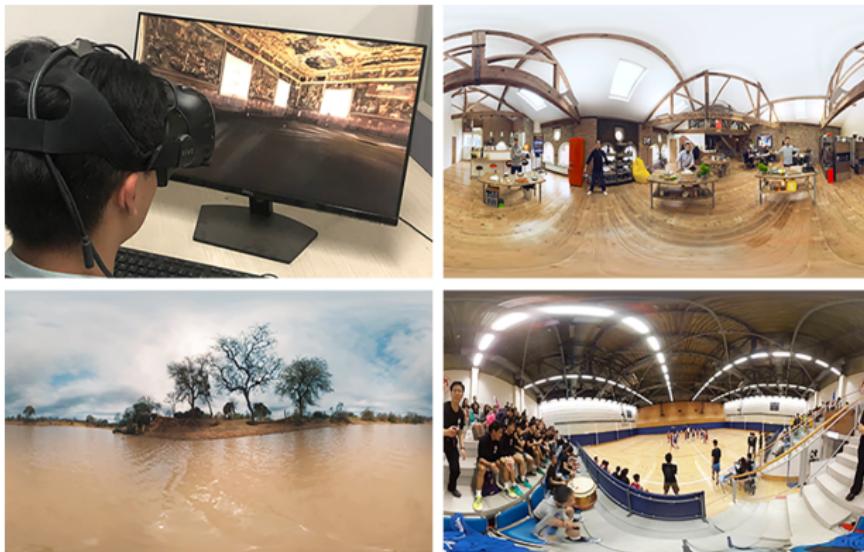
Gaze-based interaction [Mardanbegi IEEE VR'19]

Applications of human eye gaze in XR



Gaze-based visual element optimisation [Alghofaili IEEE VR'19]

Applications of human eye gaze in XR



Gaze-based activity recognition [Hu TVCG'22]

Table of Contents

Research Background

Related Work

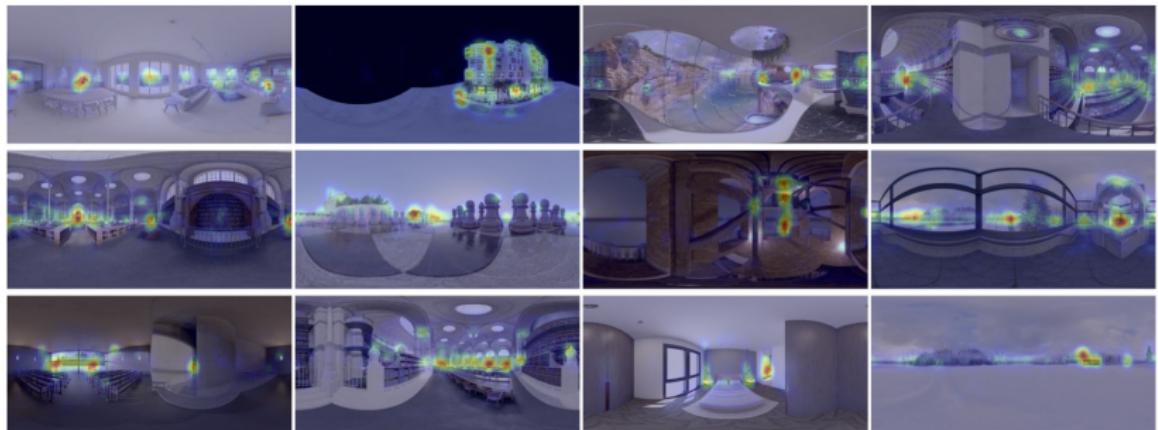
Coordination of Eye Gaze and Full-Body Poses

Coordination of Eye Gaze and Hand Gestures

Discussion

Conclusion

Saliency prediction for 360-degree images



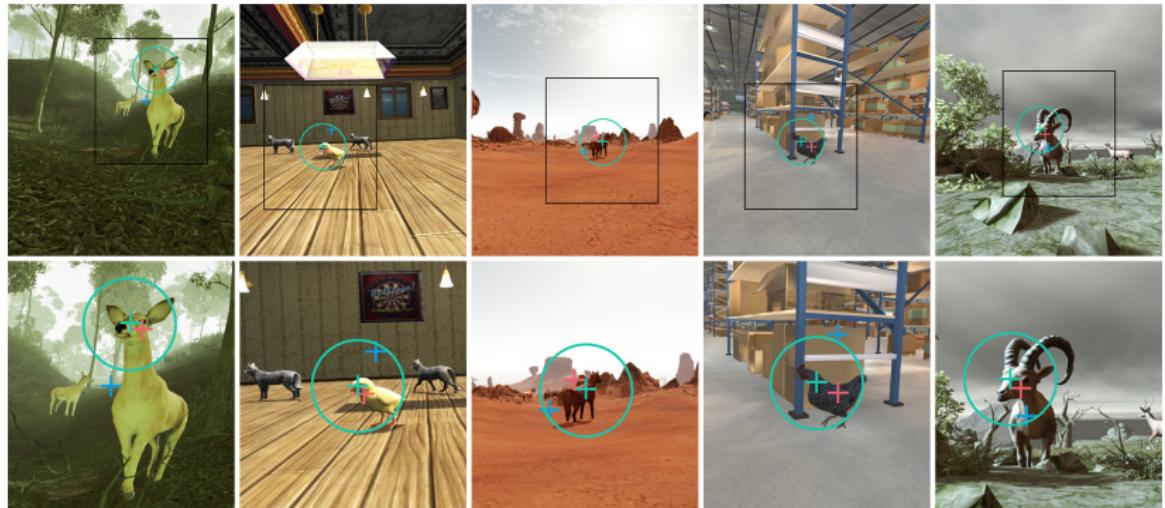
[Sitzmann TVCG'18]

Gaze analysis and estimation in static virtual environments



[Hu TCG'19]

Gaze analysis and estimation in dynamic virtual environments



[Hu TVCG'20]

Previous works

- Gaze analysis and estimation in **free-viewing** or **non-interactive** scenarios

Our work

- Gaze analysis and estimation during **human-object** and **human-human interactions**

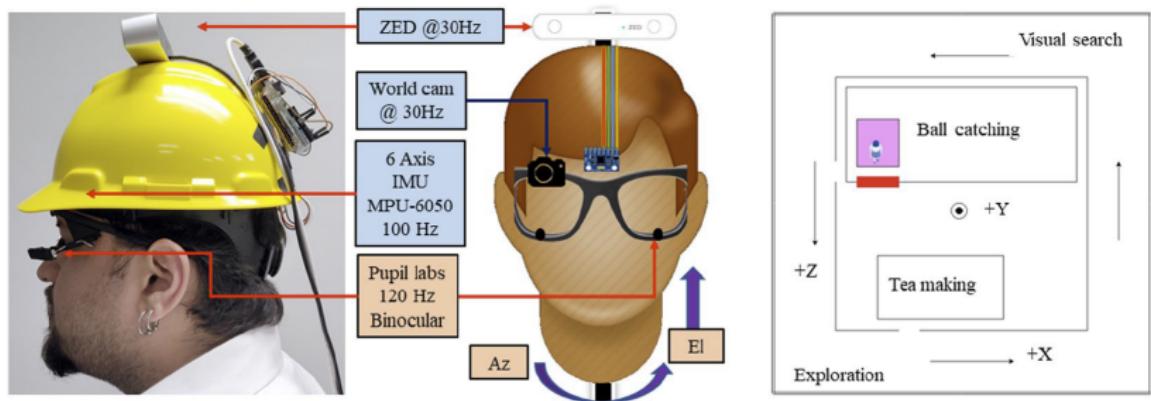
Related Work: Eye-Body Coordination

Eye-head coordination during gaze shift in XR



[Sidenmark ToCHI'19]

Eye-head coordination in everyday activities



[Kothari Sci. Rep.'20]

Eye-hand coordination during object manipulation in XR



[Belardinelli IROS'22]

Previous works

- Focus on correlations between **eye gaze** and a **specific body part**, e.g., head or hand
- Only explore correlation between eye gaze and **hand trajectories**

Our work

- Simultaneously investigate coordination of eye gaze and **full-body movements**
- Focus on coordination of eye gaze and **hand gestures**

Table of Contents

Research Background

Related Work

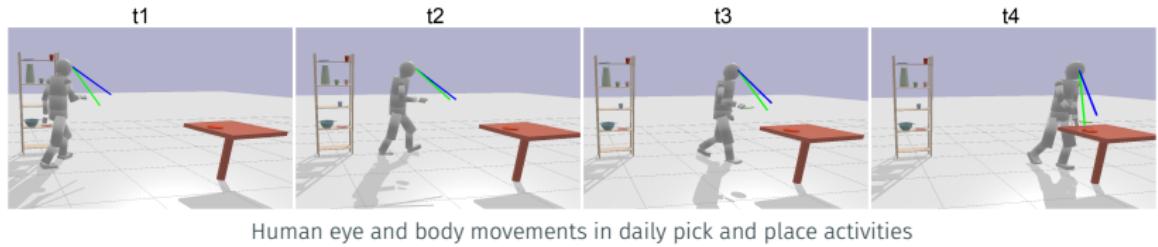
Coordination of Eye Gaze and Full-Body Poses

Coordination of Eye Gaze and Hand Gestures

Discussion

Conclusion

Eye and body movements are coordinated in daily activities



Explore eye-body coordination and predict eye gaze from full-body poses

Datasets

- **MoGaze** [Kratzer RAL'20]: real-world human-object interactions
- **ADT** [Pan ICCV'23]: XR human-object interactions
- **GIMO** [Zheng ECCV'22]: XR human-object interactions
- **EgoBody** [Zhang ECCV'22]: XR human-human interactions

Correlations between eye gaze and body orientations

		<i>base</i>	<i>pelvis</i>	<i>torso</i>	<i>neck</i>	<i>head</i>
<i>MoGaze</i>	<i>pick</i>	0.64	0.60	0.66	0.84	0.92
	<i>place</i>	0.62	0.58	0.63	0.84	0.92
<i>GIMO</i>	<i>change</i>	0.76	0.86	0.86	0.90	0.93
	<i>interact</i>	0.72	0.82	0.83	0.87	0.93
	<i>rest</i>	0.67	0.82	0.83	0.87	0.92
<i>EgoBody</i>	<i>catch</i>	0.90	0.94	0.94	0.96	0.97
	<i>chat</i>	0.81	0.85	0.87	0.90	0.94
	<i>dance</i>	0.82	0.86	0.87	0.93	0.97
	<i>discuss</i>	0.88	0.88	0.91	0.93	0.94
	<i>learn</i>	0.70	0.75	0.77	0.84	0.89
	<i>perform</i>	0.90	0.92	0.92	0.95	0.97
	<i>teach</i>	0.84	0.84	0.86	0.89	0.93

Gaze direction is strongly correlated with body orientations, especially with head direction

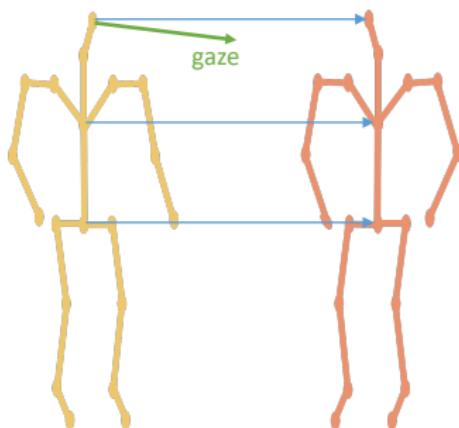
Analysis of Eye-Body Coordination

Correlations between eye gaze and body motions

		base	pelvis	torso	neck	head	l_col	r_col	l_sho	r_sho	l_elb	r_elb	l_wri	r_wri	l_hip	r_hip	l_kne	r_kne	l_ank	r_ank	l_toe	r_toe	Average
MoGaze	pick	0.40	0.40	0.41	0.42	0.46	0.42	0.42	0.38	0.41	0.35	0.40	0.34	0.46	0.40	0.40	0.42	0.42	0.31	0.32	0.37	0.37	0.39
	place	0.48	0.49	0.49	0.50	0.54	0.50	0.50	0.47	0.48	0.44	0.47	0.43	0.58	0.49	0.48	0.50	0.50	0.39	0.39	0.45	0.45	0.48
ADT	decoration	0.28	0.28	0.26	0.26	0.27	0.26	0.26	0.26	0.24	0.27	0.23	0.31	0.25	0.28	0.28	0.26	0.15	0.12	0.20	0.14	0.25	
	meal	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.19	0.19	0.20	0.22	0.22	0.20	0.20	0.20	0.20	0.09	0.08	0.13	0.10	0.18
	work	0.18	0.19	0.19	0.20	0.22	0.20	0.20	0.20	0.18	0.19	0.17	0.20	0.18	0.18	0.18	0.18	0.18	0.10	0.09	0.14	0.10	0.17
GIMO	change	0.34	0.34	0.35	0.35	0.34	0.35	0.35	0.34	0.34	0.33	0.33	0.29	0.32	0.34	0.34	0.31	0.31	0.19	0.17	0.15	0.10	0.30
	interact	0.38	0.38	0.37	0.37	0.36	0.37	0.37	0.36	0.36	0.35	0.36	0.32	0.36	0.38	0.38	0.35	0.34	0.21	0.21	0.18	0.15	0.33
	rest	0.36	0.35	0.35	0.34	0.34	0.35	0.35	0.34	0.34	0.32	0.32	0.30	0.32	0.36	0.37	0.33	0.33	0.20	0.18	0.17	0.14	0.31
EgoBody	catch	0.03	0.02	0.02	0.01	0.02	0.02	0.01	0.03	0.00	0.03	-0.02	0.04	0.00	0.03	0.02	0.02	0.02	0.02	0.00	0.02	0.01	0.02
	chat	0.01	0.01	0.01	0.02	0.02	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.02	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.01	0.01
	dance	0.05	0.05	0.05	0.04	0.04	0.04	0.05	0.04	0.04	0.03	0.04	0.03	0.03	0.05	0.05	0.05	0.04	0.02	0.01	0.02	0.02	0.04
	discuss	0.02	0.02	0.03	0.03	0.04	0.03	0.03	0.04	0.03	0.02	0.03	0.03	0.03	0.01	0.01	0.00	0.02	0.00	0.00	0.01	0.01	0.02
	learn	0.00	-0.01	-0.01	-0.01	-0.01	-0.01	0.00	-0.01	0.00	-0.01	-0.01	0.00	0.00	0.00	0.00	-0.01	-0.01	0.00	0.01	0.01	0.01	0.00
	perform	0.04	0.04	0.02	0.02	0.01	0.01	0.03	-0.01	0.01	-0.03	0.01	0.01	0.02	0.04	0.05	0.03	0.01	0.01	0.03	0.02	0.03	0.02
	teach	0.00	0.00	0.01	0.01	0.02	0.01	0.01	0.00	0.02	0.00	0.02	0.01	0.02	0.00	0.01	0.01	0.02	0.02	0.02	0.02	0.02	0.01

Eye gaze has strong correlations with body motions in **human-object** interaction activities while having little or no correlation in **human-human** interactions

Eye-body coordination in human-human interactions



Eye gaze and the directions pointing from a person's body to the body of the interaction partner

Analysis of Eye-Body Coordination

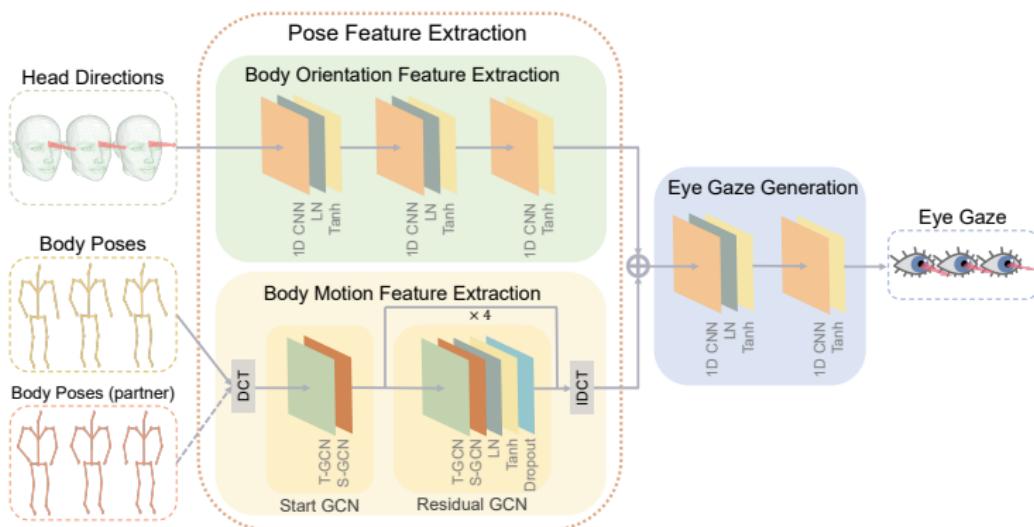
Eye-body coordination in human-human interactions

	base	pelvis	torso	neck	head	l_col	r_col	l_sho	r_sho	l_elb	r_elb	l_wri	r_wri	l_hip	r_hip	l_kne	r_kne	l_ank	r_ank	l_toe	r_toe	Average
EgoBody	catch	0.92	0.92	0.92	0.92	0.91	0.92	0.92	0.91	0.91	0.90	0.90	0.89	0.89	0.92	0.92	0.91	0.91	0.92	0.91	0.90	0.91
	chat	0.94	0.94	0.94	0.94	0.94	0.94	0.94	0.92	0.93	0.91	0.91	0.89	0.89	0.93	0.93	0.91	0.92	0.91	0.91	0.89	0.92
	dance	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.94	0.94	0.91	0.92	0.87	0.90	0.94	0.95	0.92	0.93	0.91	0.93	0.89	0.92
	discuss	0.93	0.93	0.93	0.93	0.94	0.93	0.93	0.93	0.92	0.92	0.90	0.91	0.88	0.93	0.93	0.92	0.92	0.92	0.92	0.91	0.92
	learn	0.93	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.92	0.90	0.91	0.87	0.91	0.92	0.92	0.91	0.92	0.92	0.89	0.91	0.91
	perform	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.97	0.96	0.96	0.95	0.95	0.94	0.97	0.97	0.96	0.95	0.96	0.95	0.96	0.96
	teach	0.93	0.93	0.93	0.93	0.93	0.93	0.92	0.93	0.91	0.92	0.90	0.91	0.93	0.93	0.92	0.93	0.92	0.92	0.91	0.92	0.92

Eye gaze is highly correlated with the directions between two bodies

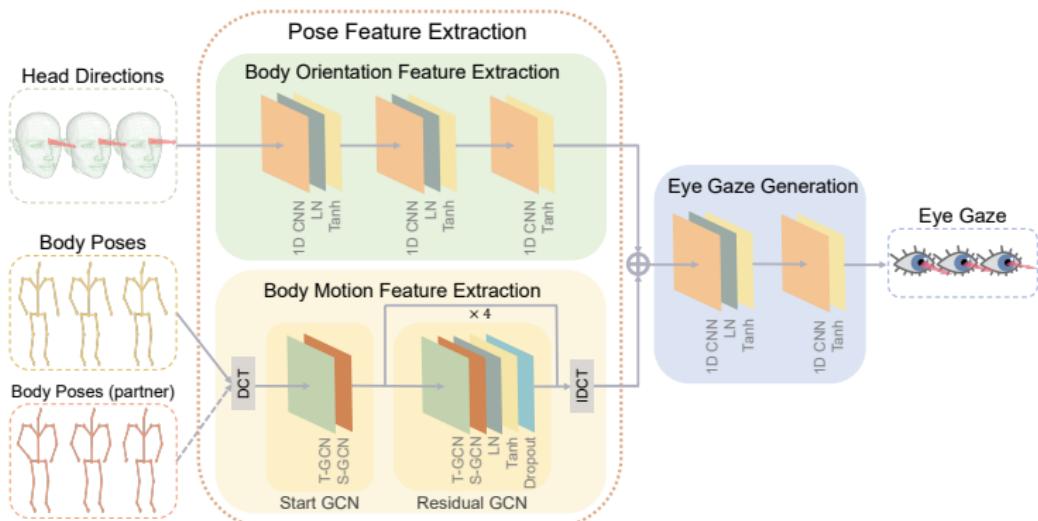
Pose2Gaze method: Body orientation feature extraction

- Use head directions as input
- 1D convolutional neural network



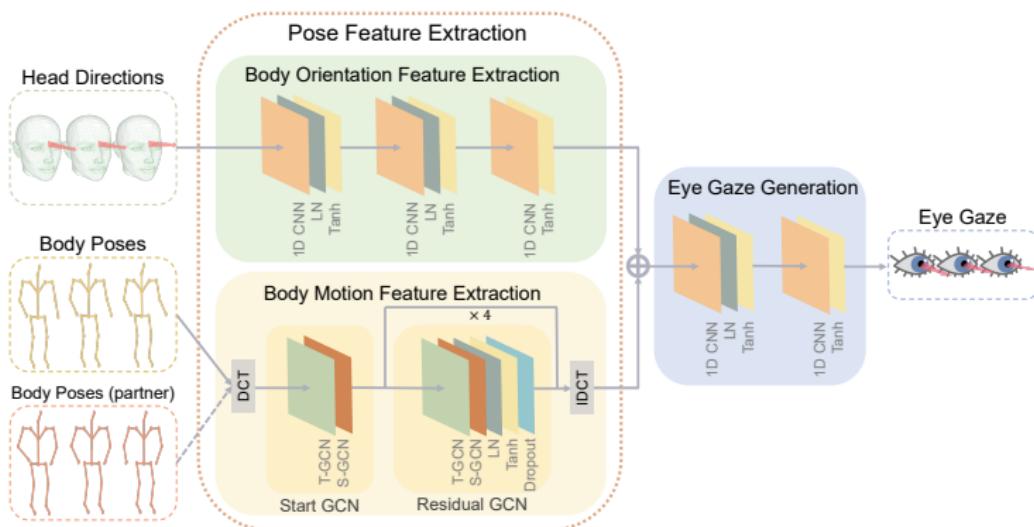
Pose2Gaze method: Body motion feature extraction

- Use body poses as input in human-object interactions
- Add partner's poses as input in human-human interactions
- Spatio-temporal graph convolutional network



Pose2Gaze method: Eye gaze generation

- Concatenate body orientation and motion features
- 1D convolutional neural network



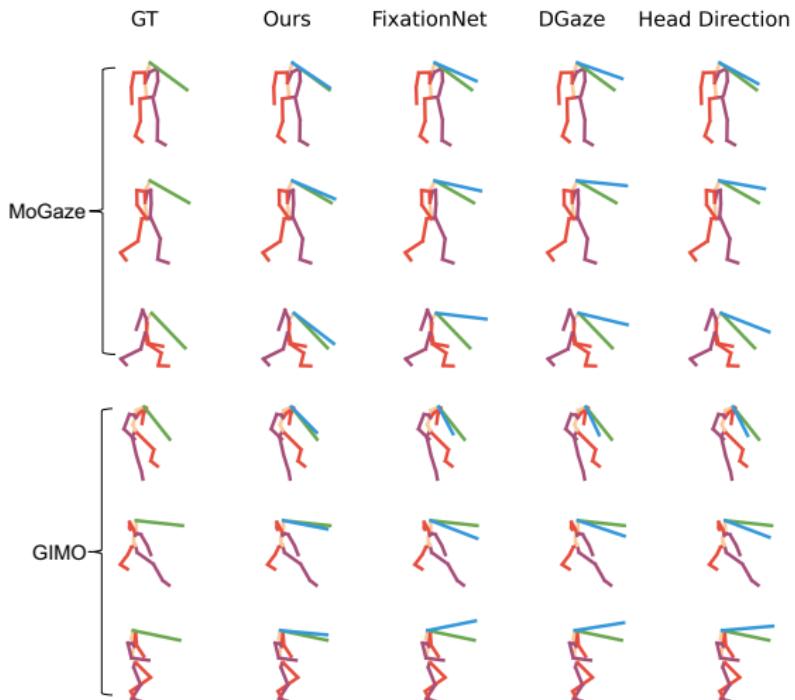
Results

Gaze estimation performance

	MoGaze			ADT				GIMO				EgoBody							
	pick	place	All	decoration	meal	work	All	change	interact	rest	All	catch	chat	dance	discuss	learn	perform	teach	All
Head Direction	17.6°	16.2°	16.9°	18.5°	25.3°	22.9°	22.3°	20.9°	19.9°	18.6°	19.8°	12.4°	16.8°	19.0°	16.6°	16.6°	14.3°	23.7°	17.7°
DGaze [Hu TVCG'20]	13.4°	12.1°	12.8°	10.3°	10.8°	8.8°	9.9°	22.6°	20.5°	17.3°	20.2°	14.1°	16.5°	22.0°	16.3°	14.8°	17.4°	24.1°	17.5°
FixationNet [Hu TVCG'21]	13.2°	11.7°	12.5°	11.2°	11.7°	9.5°	10.6°	21.7°	19.6°	17.5°	19.7°	13.9°	16.3°	21.8°	16.1°	15.1°	17.2°	23.7°	17.3°
Ours	10.7°	9.4°	10.1°	9.5°	9.8°	8.1°	9.0°	15.9°	17.3°	15.9°	16.3°	12.1°	13.5°	16.7°	14.2°	9.7°	12.0°	13.0°	13.0°

Our method significantly outperforms prior methods for both human-object and human-human interactions

Gaze estimation performance



Ablation study

	Ours	w/o DCT	w/o S-GCN	w/o T-GCN	w/o Pose	w/o Pose_I	w/o Head
ADT	9.0°	9.1°	9.4°	9.1°	9.5°	-	17.7°
GIMO	16.3°	17.3°	18.1°	17.3°	20.8°	-	20.9°
EgoBody	13.0°	13.1°	14.3°	13.3°	18.1°	17.9°	14.5°

Our method **consistently outperforms** the ablated versions

Table of Contents

Research Background

Related Work

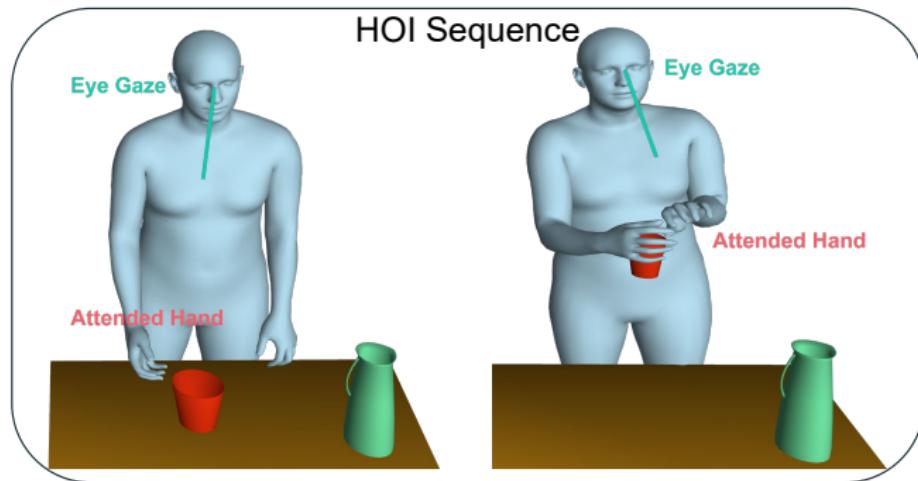
Coordination of Eye Gaze and Full-Body Poses

Coordination of Eye Gaze and Hand Gestures

Discussion

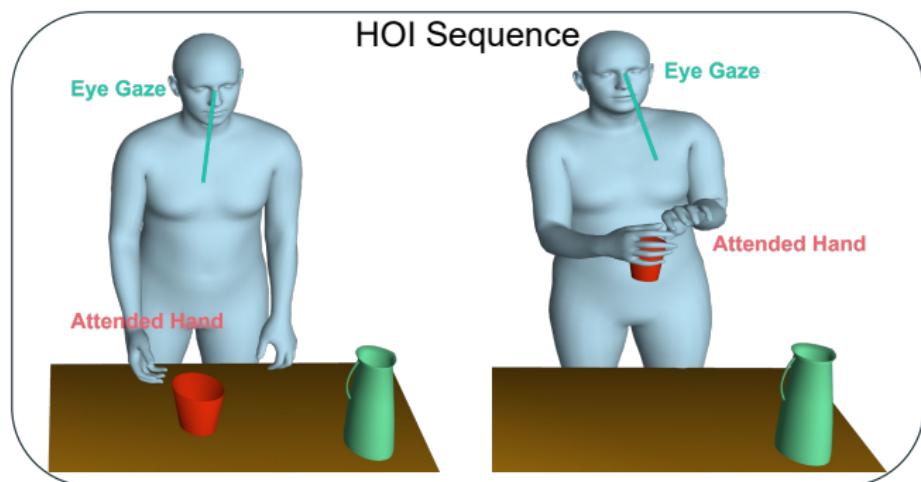
Conclusion

Eye-hand-head coordination during hand-object interactions

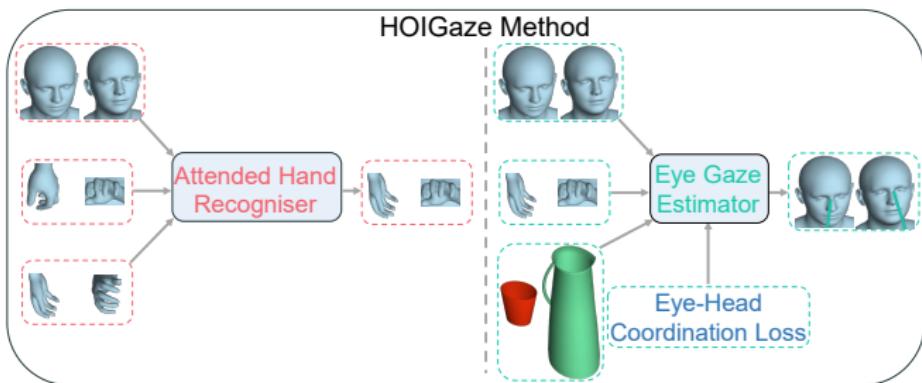


Estimate eye gaze during HOIs using eye-hand-head coordination

Attended hand: the hand that is closer to eye gaze in terms of angular distance

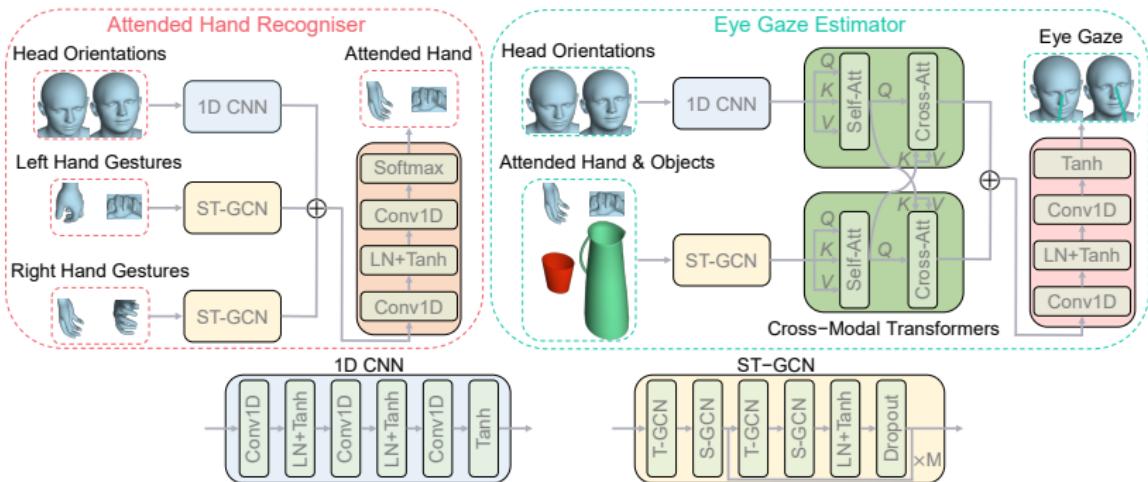


A novel hierarchical framework that first **recognises attended hand** and then **estimates eye gaze** based on the attended hand



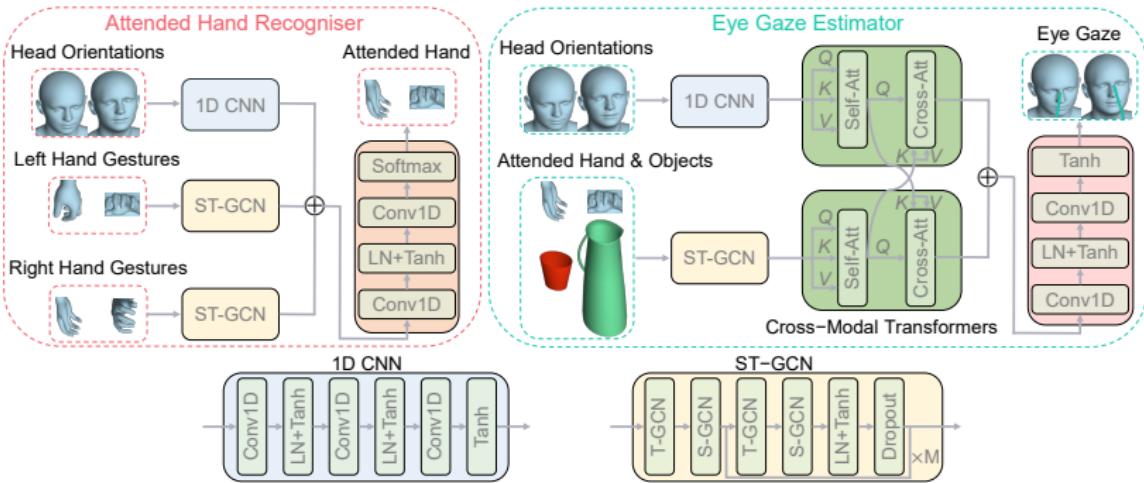
Attended hand recogniser

- 1D CNN for head orientations
- Spatio-temporal GCN for hand gestures



Eye gaze estimator

- 1D CNN for head orientations
- Spatio-temporal GCN for hand gestures and scene objects
- Cross-modal transformers for fusing features



Eye-head coordination loss that increases the weights of the loss assigned to eye-head-coordinated samples:

$$\ell_i = \begin{cases} f_{eh} * (g_i - \hat{g}_i)^2, & \text{if } g_i \cdot h_i > \text{Cos}_{eh} \\ (g_i - \hat{g}_i)^2, & \text{otherwise} \end{cases}$$

Insight: eye-head coordinated samples are **more important** than the samples with little eye-head correlation

Evaluation settings

- Datasets: **HOT3D** [Banerjee arXiv'24] for HOI setting and **ADT** [Pan ICCV'23] for general setting
- Metric: mean angular error
- Input: 15 frames of hand-head-object data
- Output: 15 frames of eye gaze directions

Results

Gaze estimation performance

	HOT3D (Cross-User)				HOT3D (Cross-Scene)				ADT			
	{P1, P2, P3}	{P9, P10, P11}	{P12, P14, P15}	Average	Room	Kitchen	Office	Average	Work	Decoration	Meal	Average
Head Direction	23.24°	28.00°	17.85°	23.20°	23.69°	22.83°	23.16°	23.20°	22.88°	18.44°	25.23°	22.25°
DGaze [Hu TVCG'20]	12.17°	15.08°	14.87°	14.29°	13.37°	12.98°	11.39°	12.81°	8.84°	10.53°	10.77°	9.92°
FixationNet [Hu TVCG'21]	11.90°	14.60°	14.78°	14.00°	12.78°	12.84°	11.34°	12.53°	8.82°	10.50°	10.83°	9.92°
Pose2Gaze [Hu TVCG'24]	10.69°	10.73°	11.80°	11.10°	9.79°	9.73°	9.96°	9.80°	8.25°	9.71°	10.43°	9.34°
Ours	9.23°	9.16°	9.69°	9.37°	8.55°	8.69°	8.69°	8.64°	7.81°	9.46°	9.41°	8.78°

Our method **significantly outperforms** prior methods for both cross-user and cross-scene evaluations

Results

Gaze estimation performance



Our method achieves **good estimation accuracy** during HOIs

Results

Ablation study

	HOT3D (Cross-User)				HOT3D (Cross-Scene)				ADT			
	{P1, P2, P3}	{P9, P10, P11}	{P12, P14, P15}	Average	Room	Kitchen	Office	Average	Work	Decoration	Meal	Average
Ours w/o attended hand	9.89°	11.24°	10.57°	10.67°	9.71°	9.32°	9.16°	9.43°	8.26°	9.97°	9.87°	9.25°
Ours w/o Transformers	9.60°	10.07°	10.24°	10.02°	8.87°	8.85°	9.17°	8.92°	8.03°	9.74°	9.96°	9.12°
Ours w/o eye-head coord. loss	9.83°	9.48°	9.70°	9.64°	8.79°	8.71°	8.84°	8.76°	7.87°	9.49°	9.71°	8.90°
Ours	9.23°	9.16°	9.69°	9.37°	8.55°	8.69°	8.69°	8.64°	7.81°	9.46°	9.41°	8.78°

Each component contributes to our method's performance

Table of Contents

Research Background

Related Work

Coordination of Eye Gaze and Full-Body Poses

Coordination of Eye Gaze and Hand Gestures

Discussion

Conclusion

Limitations

- Ignore the influence of **image/texture features** on eye gaze
- Evaluation datasets only contain interactions with **real physical objects**

Future work

- Integrate **image/texture features** into our method to further boost its performance
- Evaluate on interactions with **both real and virtual objects**
- Explore eye-body coordination for interactions between **more humans** or between **a human and a virtual avatar**

Table of Contents

Research Background

Related Work

Coordination of Eye Gaze and Full-Body Poses

Coordination of Eye Gaze and Hand Gestures

Discussion

Conclusion

Main contributions

- A comprehensive analysis of **eye-body coordination** in **human-object** and **human-human** interaction activities
- A thorough investigation of the correlation between **eye gaze** and **hand gestures** during **hand-object** interactions
- Two novel **gaze estimation models** that exploit eye-hand-body coordination and achieve **state-of-the-art** performances



References i

- Alghofaili IEEE VR'19. Optimizing visual element placement via visual attention analysis. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces*, pages 464–473. IEEE, 2019.
- Banerjee arXiv'24. Introducing hot3d: An egocentric dataset for 3d hand and object tracking. *arXiv preprint arXiv:2406.09598*, 2024.
- Belardinelli IROS'22. Intention estimation from gaze and motion features for human-robot shared-control object manipulation. In *Proceedings of the 2022 IEEE International Conference on Intelligent Robots and Systems*, pages 9806–9813. IEEE, 2022.
- 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.
- Hu TVCG'24. Pose2gaze: Eye-body coordination during daily activities for gaze prediction from full-body poses. *IEEE Transactions on Visualization and Computer Graphics*, 2024.
- Kothari Sci. Rep.'20. Gaze-in-wild: a dataset for studying eye and head coordination in everyday activities. *Scientific Reports*, 10(1):1–18, 2020.
- Kratzer RAL'20. Mogaze: A dataset of full-body motions that includes workspace geometry and eye-gaze. *IEEE Robotics and Automation Letters*, 6(2):367–373, 2020.

References ii

- Mardanbegi IEEE VR'19. Eyesethrough: unifying tool selection and application in virtual environments. In *Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces*, pages 474–483, 2019.
- Pan ICCV'23. Aria digital twin: A new benchmark dataset for egocentric 3d machine perception. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 20133–20143, 2023.
- Sidenmark ToCHI'19. Eye, head and torso coordination during gaze shifts in virtual reality. *ACM Transactions on Computer-Human Interaction*, 27(1):1–40, 2019.
- Sitzmann TVCG'18. Saliency in vr: how do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642, 2018.
- Zhang ECCV'22. Egobody: Human body shape, motion and social interactions from head-mounted devices. In *Proceedings of the 2022 European Conference on Computer Vision*, 2022.
- Zheng ECCV'22. Gimo: Gaze-informed human motion prediction in context. In *Proceedings of the 2022 European Conference on Computer Vision*, 2022.

Thank you!

Any questions?