Pose2Gaze: Eye-body Coordination during Daily Activities for Gaze Prediction from Full-body Poses

Zhiming Hu[1], Jiahui Xu[1], Syn Schmitt[1,2], Andreas Bulling[1,2]

[1]University of Stuttgart

[2]Bionic Intelligence Tuebingen Stuttgart

Table of Contents

Research Background

Related Work

Analysis of Eye-body Coordination

Method

Results

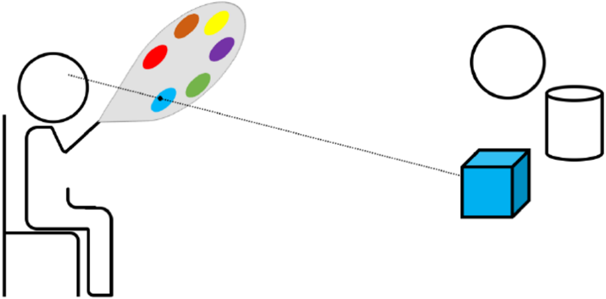
Discussion

Conclusion

Applications of human eye gaze in XR

Gaze-contingent rendering
[Hu TVCG'20]

Applications of human eye gaze in XR

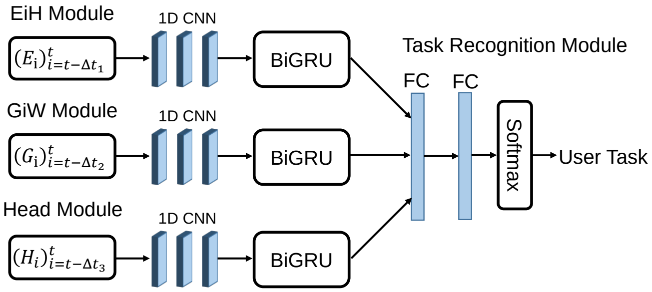


Gaze-based interaction
[Mardanbegi IEEE VR'19]

Applications of human eye gaze in XR

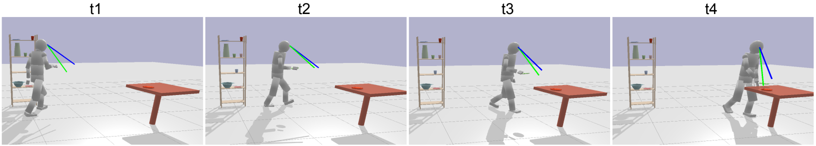
Gaze-based visual element optimisation
[Alghofaili IEEE VR'19]

Applications of human eye gaze in XR



Gaze-based activity recognition
[Hu TVCG'22]

Eye and body movements are coordinated in daily activities



Human eye and body movements in daily pick and place activities

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

Contributions

- Comprehensive **analyses** of **eye-body coordination** in diverse **human-object** and **human-human** interaction activities
- A novel method that combines a **CNN** and a **spatio-temporal GCN** to predict **eye gaze** from **full-body poses**
- Extensive experiments on **four public datasets** that demonstrate **significant improvements** over prior methods
- Experiments on the downstream task of **gaze-based activity recognition** that demonstrate our method's effectiveness

Research Background

Related Work

Analysis of Eye-body Coordination

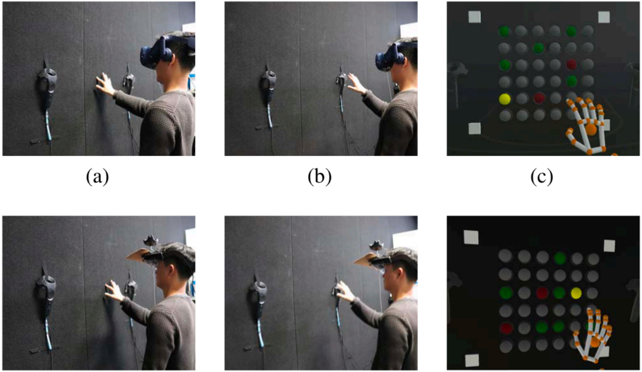
Method

Results

Discussion

Conclusion

Eye-hand coordination

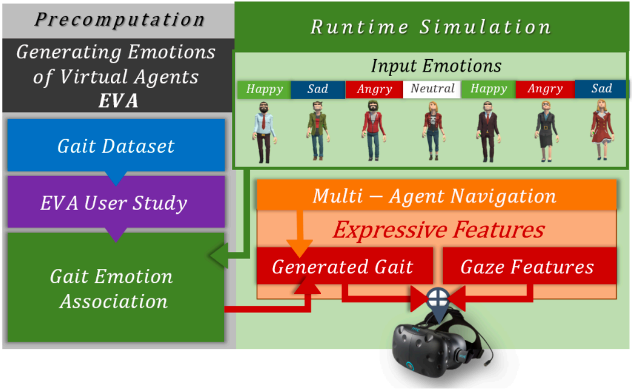


(a) (b) (c)

(d) (e) (f)

[Batmaz IEEE VR'20]

Eye-gait coordination



[Randhavane SAP'19]

Eye-head-torso coordination

[Sidenmark ToCHI'19]

Previous works

- Focus on correlations between **eye gaze** and **specific body parts** (e.g., head, hand, or torso)

Our work

- **Simultaneously** investigate coordination of **eye** and **full-body movements**

Table of Contents

Research Background

Related Work

Analysis of Eye-body Coordination

Method

Results

Discussion

Conclusion

Datasets

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

Correlations between eye gaze and body orientations

The **cosine similarities** between **eye gaze direction** and the **directions of different body joints**

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

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

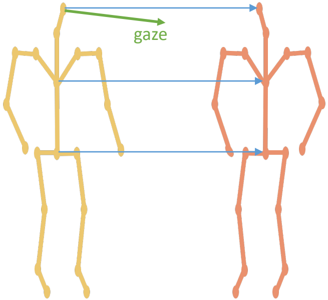
Correlations between eye gaze and body motions

The **cosine similarities** between **eye gaze** and the **motions of different body joints**

		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	0.28	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	0.19	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	-0.01	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**

Eye-body coordination in human-human interactions

The **cosine similarities** between **gaze** and the directions pointing from a person's body to the **interaction partner**

		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
	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	0.91	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	0.89	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	0.93
EgoBody	*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	0.91	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	0.91	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	0.94	0.96
	teach	0.93	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**

Research Background

Related Work

Analysis of Eye-body Coordination

Method

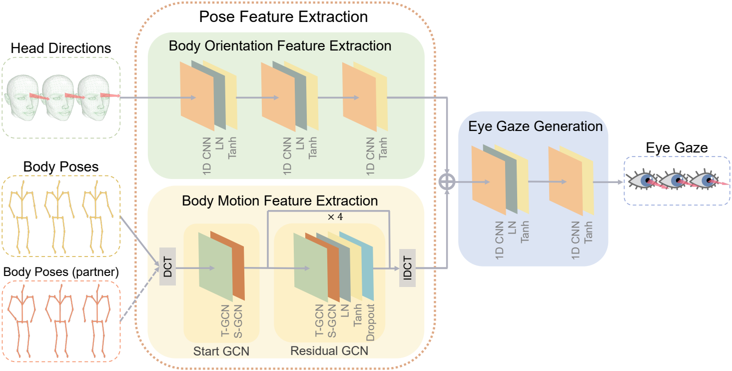
Results

Discussion

Conclusion

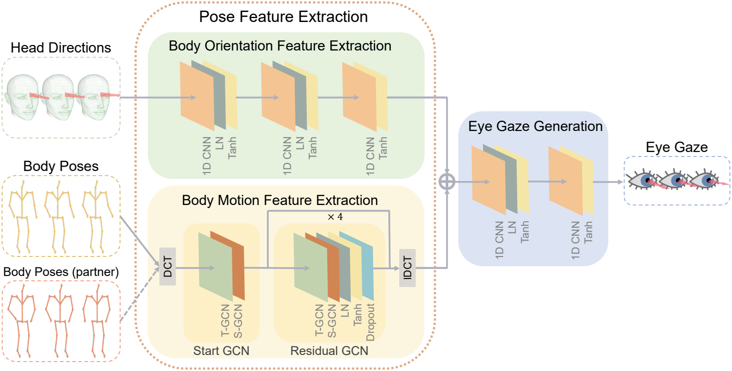
Pose2Gaze method

- Body orientation feature extraction
- Body motion feature extraction
- Eye gaze generation



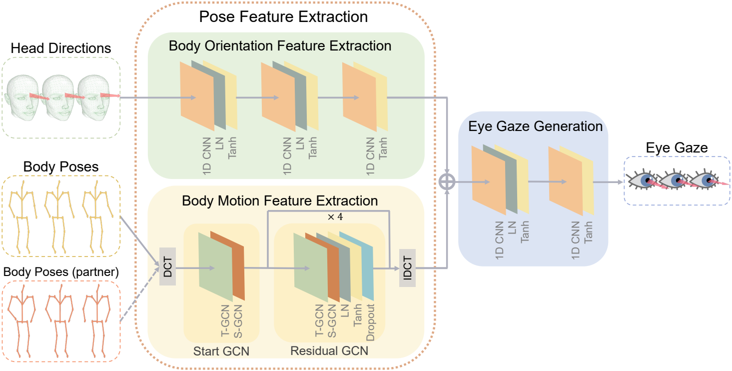
Pose2Gaze method: Body orientation feature extraction

- 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

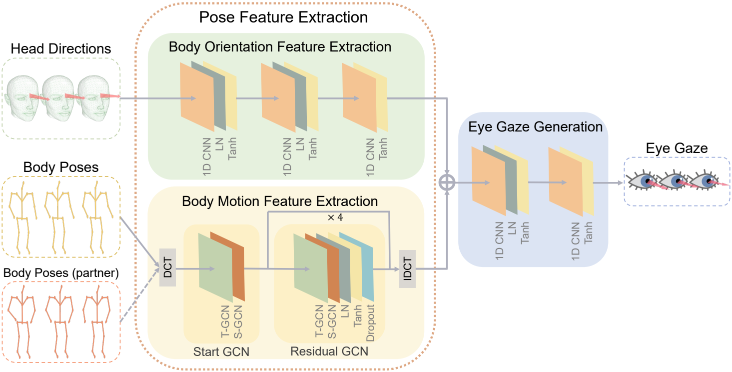


Table of Contents

Research Background

Related Work

Analysis of Eye-body Coordination

Method

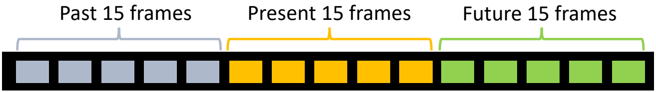
Results

Discussion

Conclusion

Eye gaze generation settings

- Generating gaze from **past** poses: **eye gaze forecasting**
- Generating gaze from **present** poses: **eye gaze real-time estimation**
- Generating gaze from **future** poses: **eye gaze offline generation**



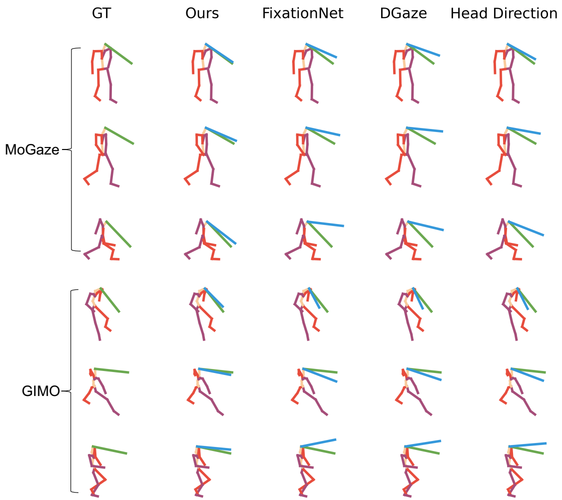
Gaze generation performance

Mean angular errors of different methods for **generating eye gaze** from **past**, **present**, and **future** body poses

		MoGaze			ADT				GIMO				EgoBody							
		pick	*place*	All	*decoration*	*meal*	*work*	All	*change*	*interact*	*rest*	All	*catch*	*chat*	*dance*	*discuss*	*learn*	*perform*	*teach*	All
past	*Head Direction*	37.8°	34.9°	36.4°	26.5°	30.6°	27.1°	28.0°	23.5°	23.7°	22.9°	23.4°	14.6°	18.1°	25.0°	18.0°	17.6°	16.8°	24.5°	19.2°
	DGaze [Hu TVCG'20]	18.3°	15.3°	16.9°	13.6°	13.2°	11.1°	12.5°	23.1°	20.4°	18.9°	20.9°	17.1°	17.9°	27.1°	19.6°	17.3°	21.0°	24.6°	19.5°
	FixationNet [Hu TVCG'21]	18.2°	15.2°	16.8°	14.8°	14.3°	12.0°	13.5°	22.2°	**20.0°**	19.7°	20.7°	15.4°	17.3°	23.7°	17.6°	16.4°	18.9°	24.5°	18.5°
	Ours	**15.0°**	**11.1°**	**13.1°**	**12.6°**	**12.2°**	**10.2°**	**11.5°**	**17.9°**	21.2°	**16.1°**	**18.4°**	**12.9°**	**13.3°**	**19.5°**	**16.0°**	**8.6°**	**13.9°**	**13.5°**	**13.2°**
present	*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°**
future	*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	**10.1°**	**8.8°**	**9.5°**	**9.7°**	**9.3°**	**7.9°**	**8.9°**	**15.4°**	**16.2°**	**14.8°**	**15.5°**	**11.1°**	**13.2°**	**15.8°**	**14.5°**	**9.2°**	**11.9°**	**13.9°**	**12.9°**

Our method **significantly outperforms** prior methods for **three different eye gaze generation** tasks

Gaze generation performance



Ablation study

Mean angular errors of different **ablated versions** of our method

		Ours	w/o *DCT*	w/o *S-GCN*	w/o *T-GCN*	w/o *Pose*	w/o *Pose_I*	w/o *Head*
ADT	*past*	**11.5°**	11.7°	11.8°	11.9°	12.2°	-	18.2°
	present	**9.0°**	9.1°	9.4°	9.1°	9.5°	-	17.7°
	future	**8.9°**	9.1°	9.3°	9.1°	9.3°	-	16.4°
GIMO	*past*	**18.4°**	19.0°	19.3°	19.1°	21.2°	-	22.1°
	present	**16.3°**	17.3°	18.1°	17.3°	20.8°	-	20.9°
	future	**15.5°**	16.6°	18.1°	16.7°	20.8°	-	18.8°
EgoBody	*past*	**13.2°**	13.5°	13.4°	13.4°	20.6°	18.6°	15.1°
	present	**13.0°**	13.1°	14.3°	13.3°	18.1°	17.9°	14.5°
	future	**12.9°**	13.7°	14.8°	13.5°	17.1°	17.9°	15.1°

Our method consistently outperforms the ablated versions

Downstream task of gaze-based activity recognition

Gaze-based activity recognition accuracies of different methods

	GT	Ours	*DGaze* [Hu TVCG'20]	*FixationNet* [Hu TVCG'21]	*Head Direction*	*Chance*
ADT	**74.7%**	<u>70.0%</u>	67.3%	66.8%	40.9%	33.3%
EgoBody	**62.1%**	<u>60.1%</u>	52.3%	58.2%	50.3%	33.3%

Our method achieves **higher** recognition accuracies than other methods and is **comparable** with the **ground truth eye gaze**

Table of Contents

Research Background

Related Work

Analysis of Eye-body Coordination

Method

Results

Discussion

Conclusion

Limitations

- Ignore the influence of the **visual scene content** on **eye-body coordination**
- Eye-body coordination analyses are limited to **indoor environments**

Future work

- Incorporate other modalities such as **facial expressions** and **audio signals** to improve gaze generation performance
- Explore eye-body coordination for interactions between **more humans** or between **a human and a virtual avatar**
- Generate **stylistic** eye gaze, e.g. eye gaze that can convey different **emotions**

Table of Contents

Research Background

Related Work

Analysis of Eye-body Coordination

Method

Results

Discussion

Conclusion

Main contributions

- **Eye-body coordination** analyses in diverse **human-object** and **human-human** interaction activities
- A novel method that employs a **CNN** and a **spatio-temporal GCN** to extract **full-body pose features** for **gaze generation**
- Extensive experiments on **four public datasets** that demonstrate the **superiority** of our method
- Experiments on the application of **gaze-based activity recognition** that validate the **effectiveness** of our method

Code available at **zhiminghu.net/hu24_pose2gaze**

Acknowledgement

Thank you!

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.

Batmaz IEEE VR'20. Touch the wall: Comparison of virtual and augmented reality with conventional 2d screen eye-hand coordination training systems. In *Proceedings of the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces*, pages 184–193. IEEE, 2020.

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.

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.

Mardanbegi IEEE VR'19. Eyeseethrough: 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.

Randhavane SAP'19. Eva: Generating emotional behavior of virtual agents using expressive features of gait and gaze. In *Proceedings of the 2019 ACM Symposium on Applied Perception*, pages 1–10, 2019.

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.

References ii

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.