

The Computer Vision Group at TUM (<https://cvg.cit.tum.de/>) is seeking two student assistants (m/f/x). In this role, you will assist with implementing and evaluating baseline methods for building egocentric world models using available egocentric video datasets [1, 2] following our previous works [3, 4]. We are looking for student assistants to support this line of research flexibly for 10-20 hours per week. In addition, we are also interested in supporting students to explore novel ideas in embodied AI, reasoning, and open-ended learning. Feel free to reach out.

We offer:

- Insight into cutting-edge research and an opportunity to collaborate with domain experts
- Practical exposure to applied computer vision
- Possibility to publish in top-tier venues as the lead author or co-author
- Flexible working hours (10-20 hours per week) & TUM-standard compensation

Responsibilities:

- Implement and run baseline methods.
- Evaluate the performance of baselines using established benchmarks.

Requirements:

- Strong knowledge of PyTorch and experience implementing neural networks
- Familiarity with 3D computer vision techniques such as 3D/4D Gaussian Splatting, Reconstruction, Multi-view geometry, etc.
- Strong communication skills and ability to work independently
- Commitment to high code quality and maintainability

The initial employment contract will be for 6 months, with the possibility of extension. The relevant project is highly research-oriented, and we encourage motivated students interested in gaining more research experience and publishing to directly contact us via email. Excellent outcomes can easily result in a publication in top-tier venues. If you're interested, please send your CV and grade reports to Abhishek Saroha at [Abhishek.Saroha@in.tum.de](mailto:Abhishek.Saroha@in.tum.de) and Xi wang at [xi.wang@inf.ethz.ch](mailto:xi.wang@inf.ethz.ch).

[1] Grauman, Kristen, Andrew Westbury, Eugene Byrne, Zachary Chavis, Antonino Furnari, Rohit Girdhar, Jackson Hamburger et al. "Ego4d: Around the world in 3,000 hours of egocentric video." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18995-19012. 2022.

[2] Grauman, Kristen, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh et al. "Ego-exo4d: Understanding skilled human activity from first-and third-person perspectives." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 19383-19400. 2024.

[3] Saroha, A., Zeng, H., Zuo, X., Cremers, D., & Wang, X. (2026). EgoFlow: Gradient-Guided Flow Matching for Egocentric 6DoF Object Motion Generation. *CVPR, 2026*.

[4] Zeng, H., Saroha, A., Cremers, D., & Wang, X. (2026). GMT: Goal-Conditioned Multimodal Transformer for 6-DOF Object Trajectory Synthesis in 3D Scenes. *3DV, 2026*.