

# Y. SATO LAB.

Connecting human perception and AI through computer vision



Department of Informatics and Electronics

Department of Information and Communication Engineering, GSIST,  
Emerging Design and Informatics Course, GSII

Computer vision

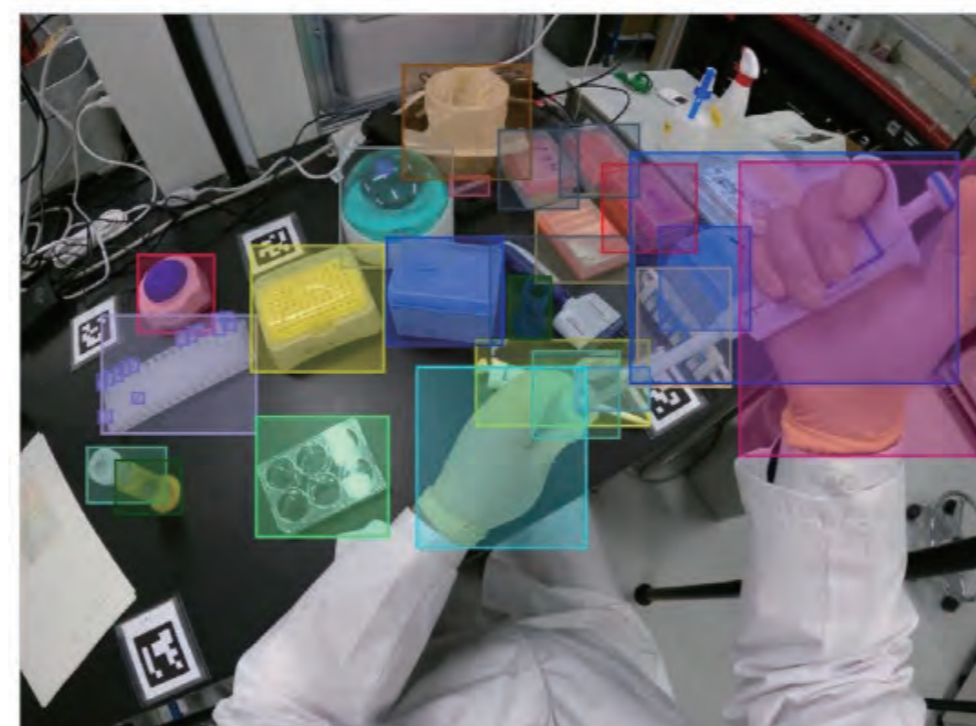
<https://www.ut-vision.org/sato-lab/>

## Paving the Future of Human-AI Co-creation via Computer Vision

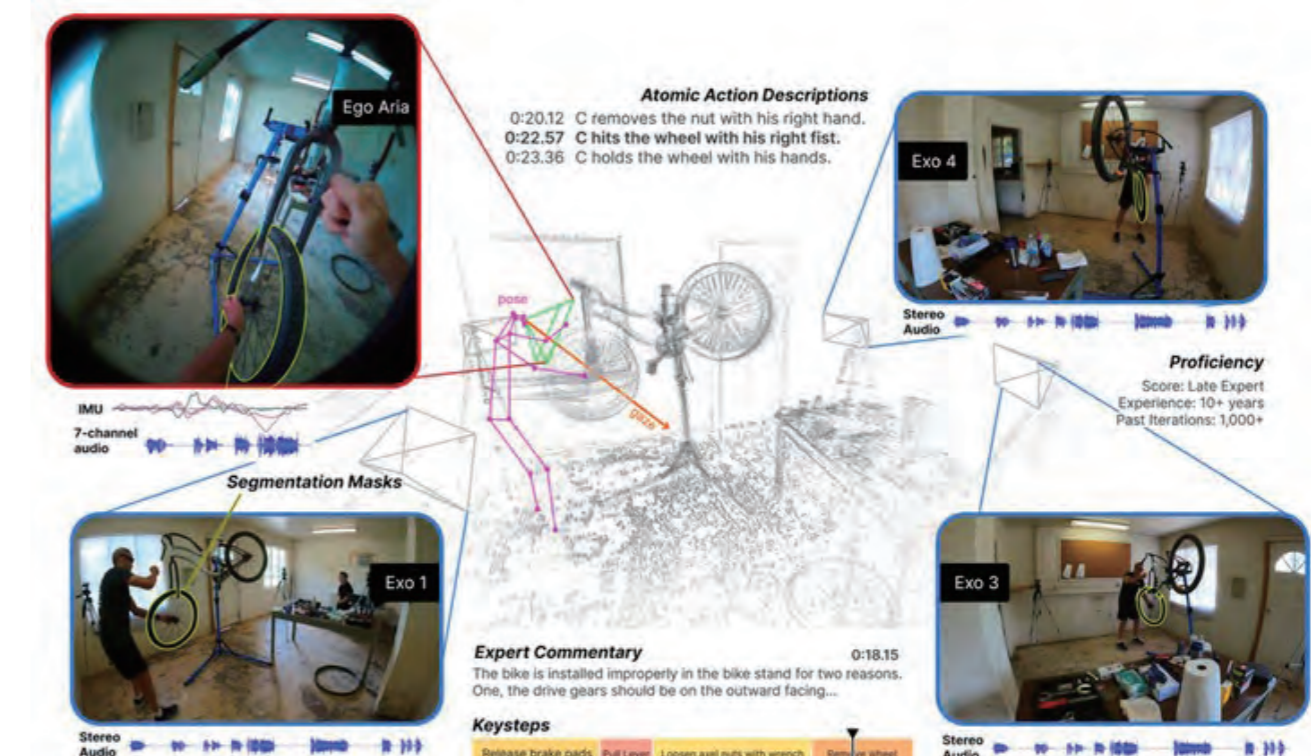
We are developing next-generation visual intelligence that achieves a deep understanding of the “what, when, how, and why” of human activities from our own first-person perspective. By integrating diverse modalities with computer vision technologies—including first-person video, gaze, motion data, and EEG—we aim at bridging human perception with artificial intelligence to realize the future of human-AI co-creation.

### Deep Understanding of Human Activities through Egocentric Vision

Going beyond action recognition, this research aims to understand human behavior from a first-person perspective. It captures how and why actions are performed, models expert skills, and analyzes team dynamics such as leadership to enable a comprehensive understanding of complex activities.



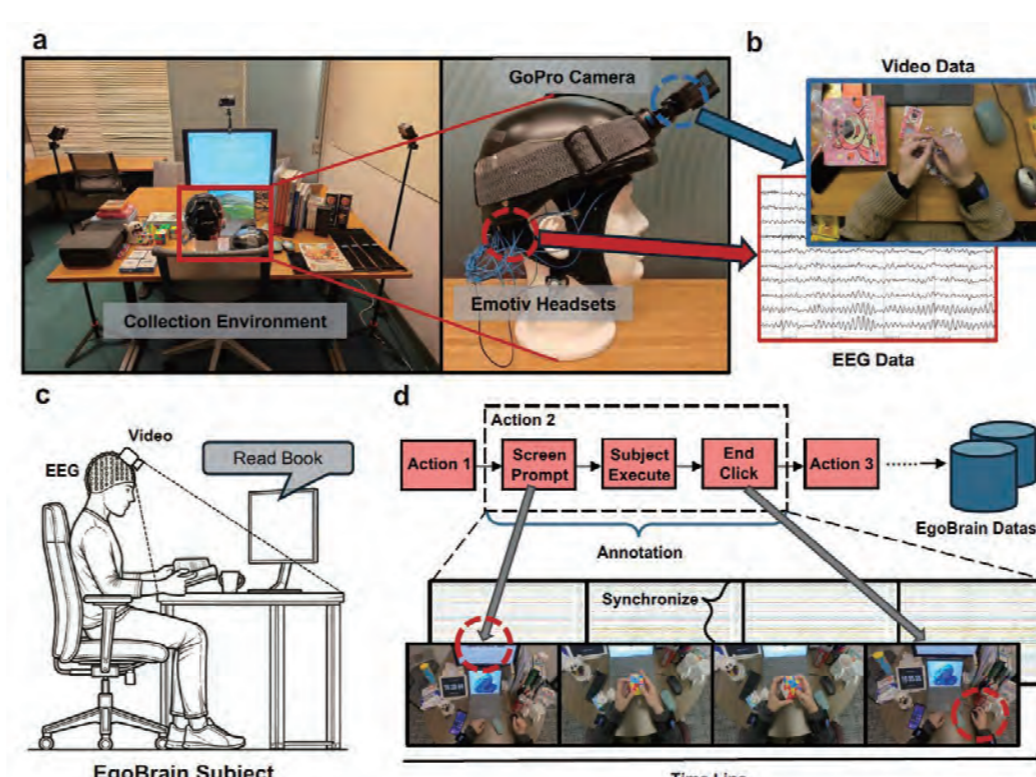
Human activity understanding in biological experiments (FineBio)



Understanding skilled activities from first-person and third-person videos (Ego-Exo4D)

### Multimodal Analysis for Understanding Human Activities and Intention

By integrating multimodal data—such as video, gaze, motion signals, and brain activity—this research enables the understanding of human behavior and underlying intent beyond what can be captured from visual information alone.



Multimodal human activity understanding with first-person videos and EEG signals (EgoBrain)

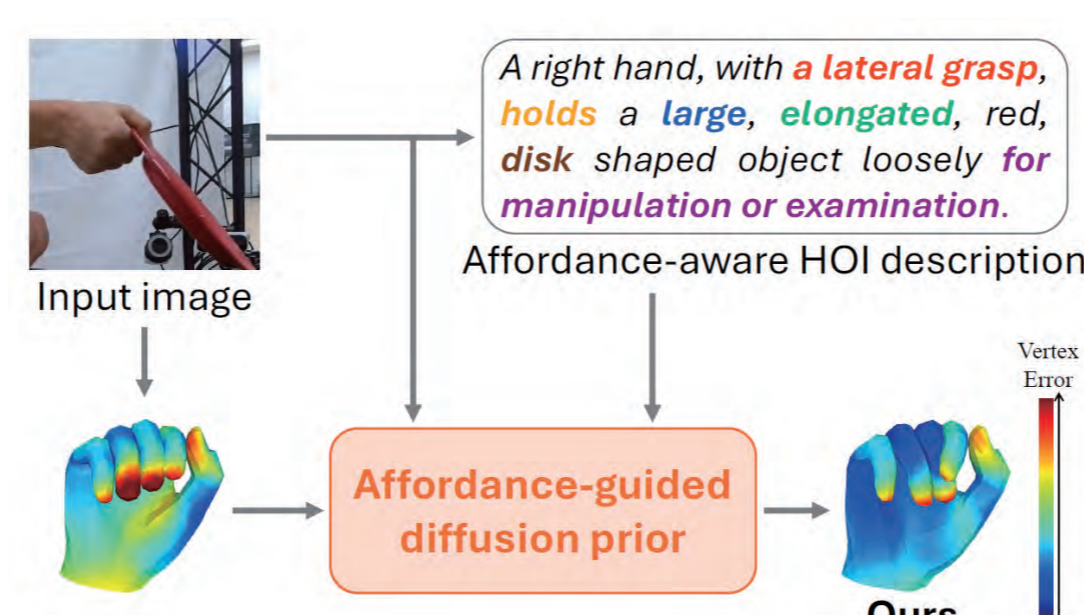
Q: Which place is shown in the picture next to the brown bear?



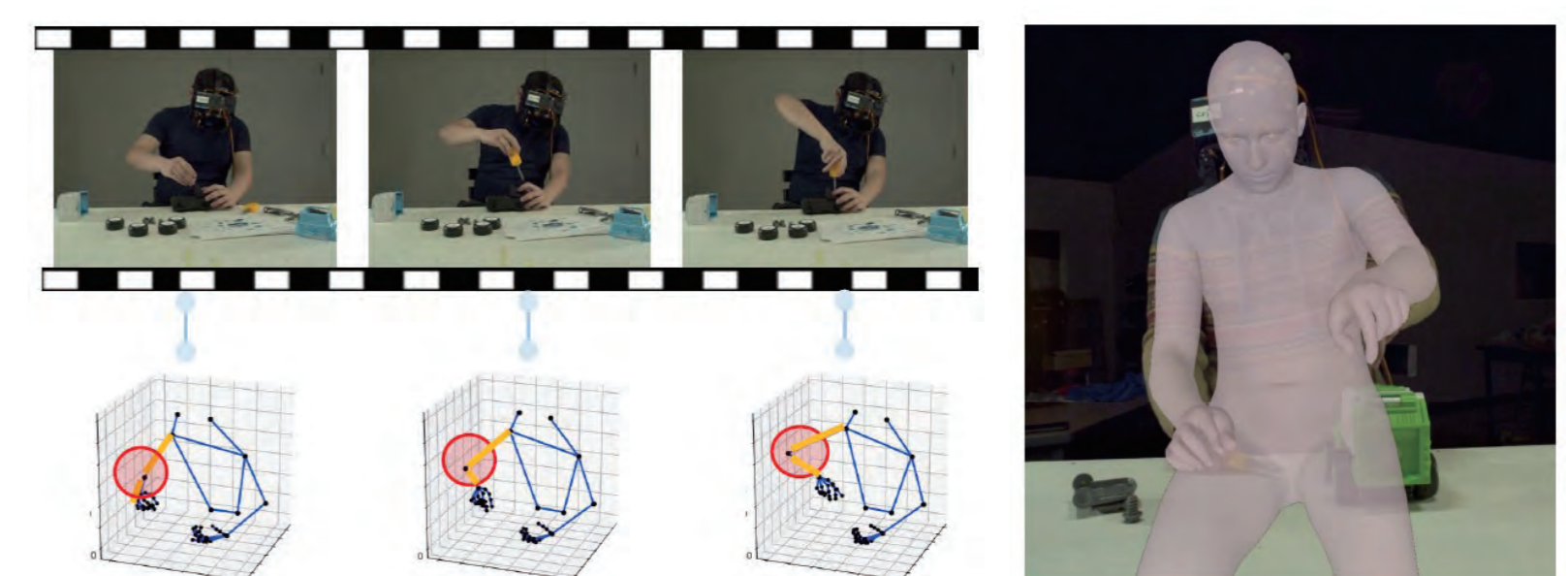
Understanding relationship between gaze behavior and underlying tasks

### Understanding Hand-Object Interactions

We develop technologies to understand how humans manipulate objects, focusing on hand-object interactions as a foundation for embodied intelligence. Our approach captures precise 3D hand pose and shape while enabling the analysis of complex object manipulation.



3D hand shape reconstruction guided by object affordances



Understanding bimanual human activities by modeling hand, body, and visual context (AssemblyHands-X)