

Active Vision Might Be All You Need

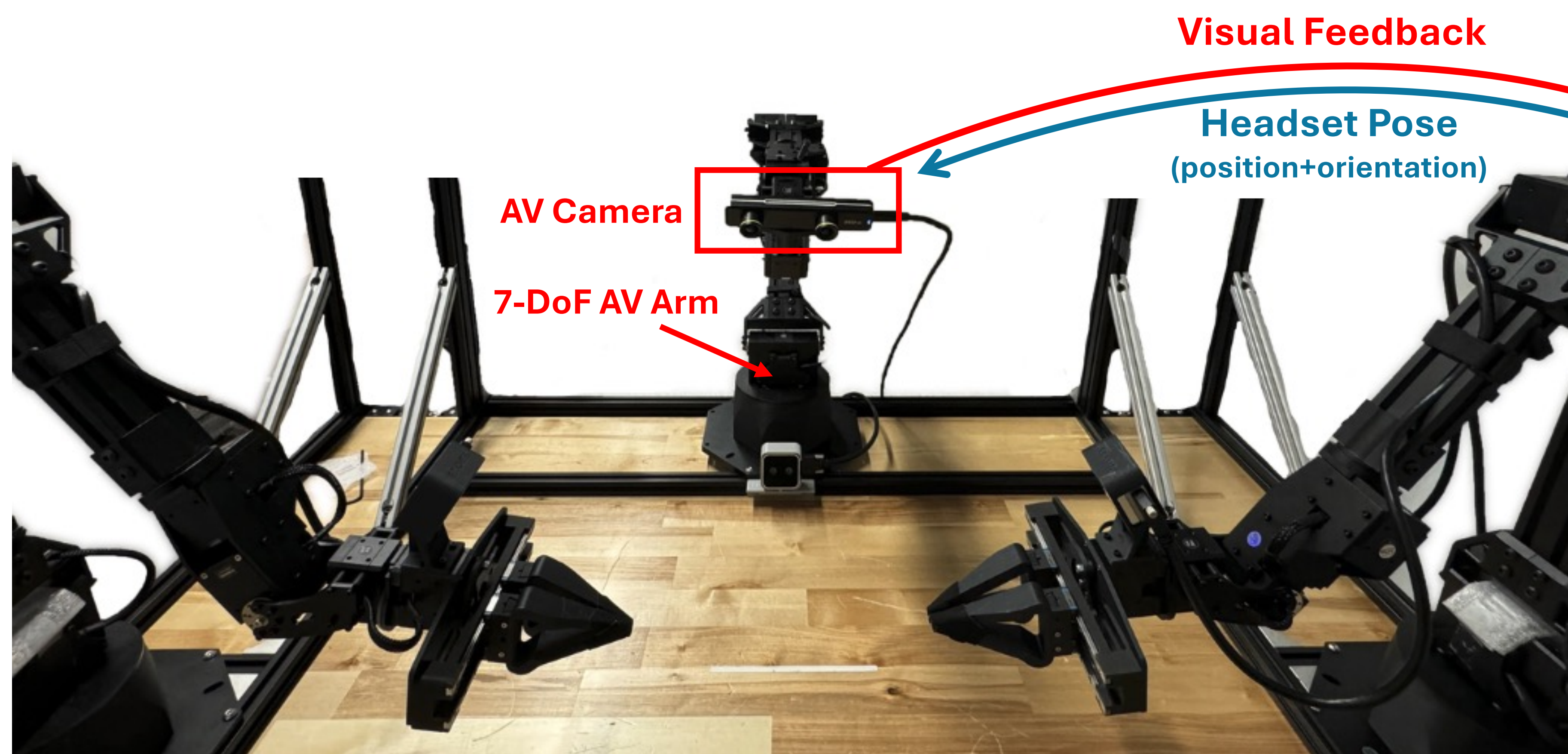
Exploring Active Vision in Bimanual Robotic Manipulation

Ian Chuang^{*1,2}, Andrew Lee^{*2}, Dechen Gao², M-Mahdi Naddaf-Sh², Iman Soltani²

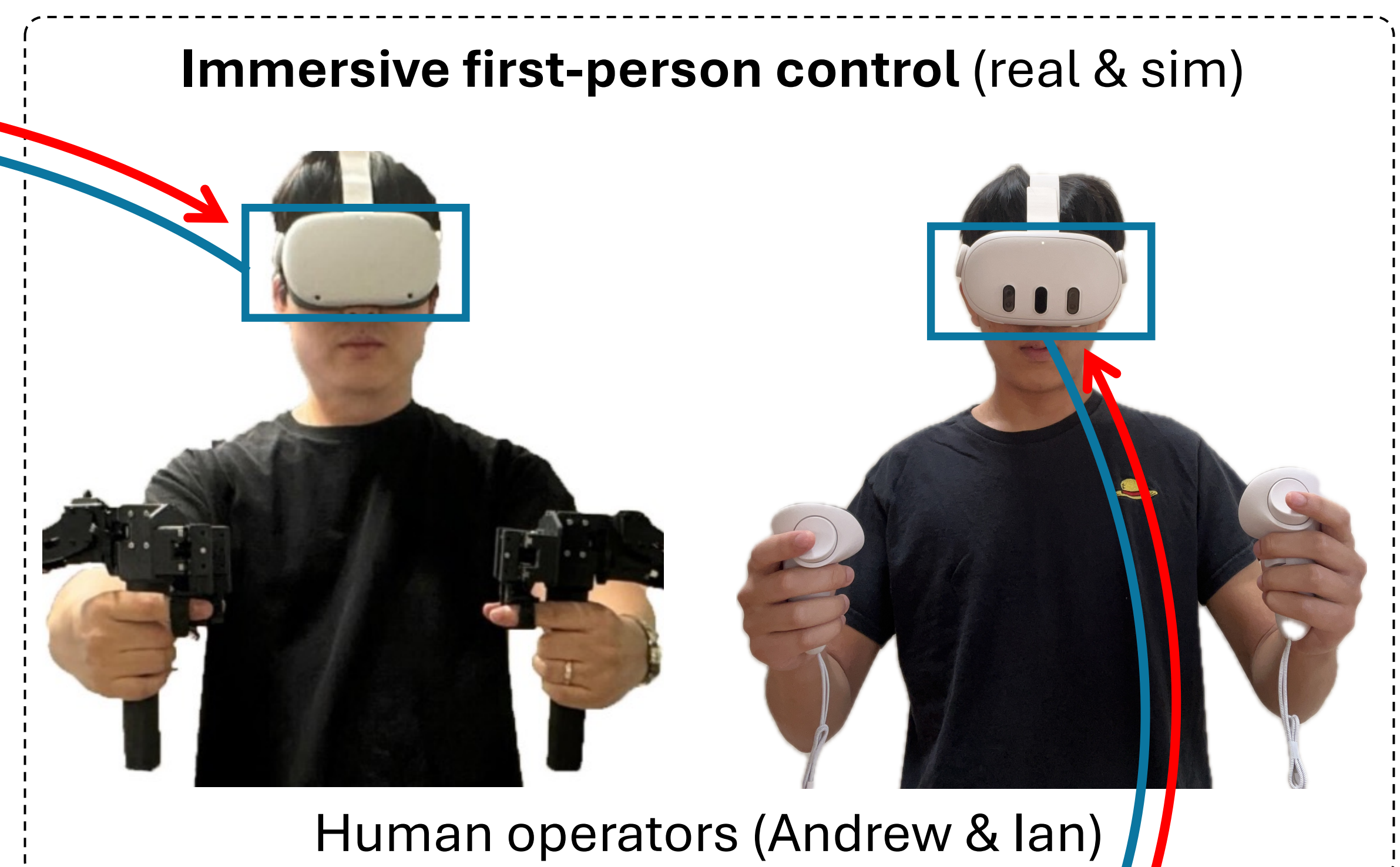
¹University of California Berkeley, ²University of California Davis ^{*}Equal Contribution



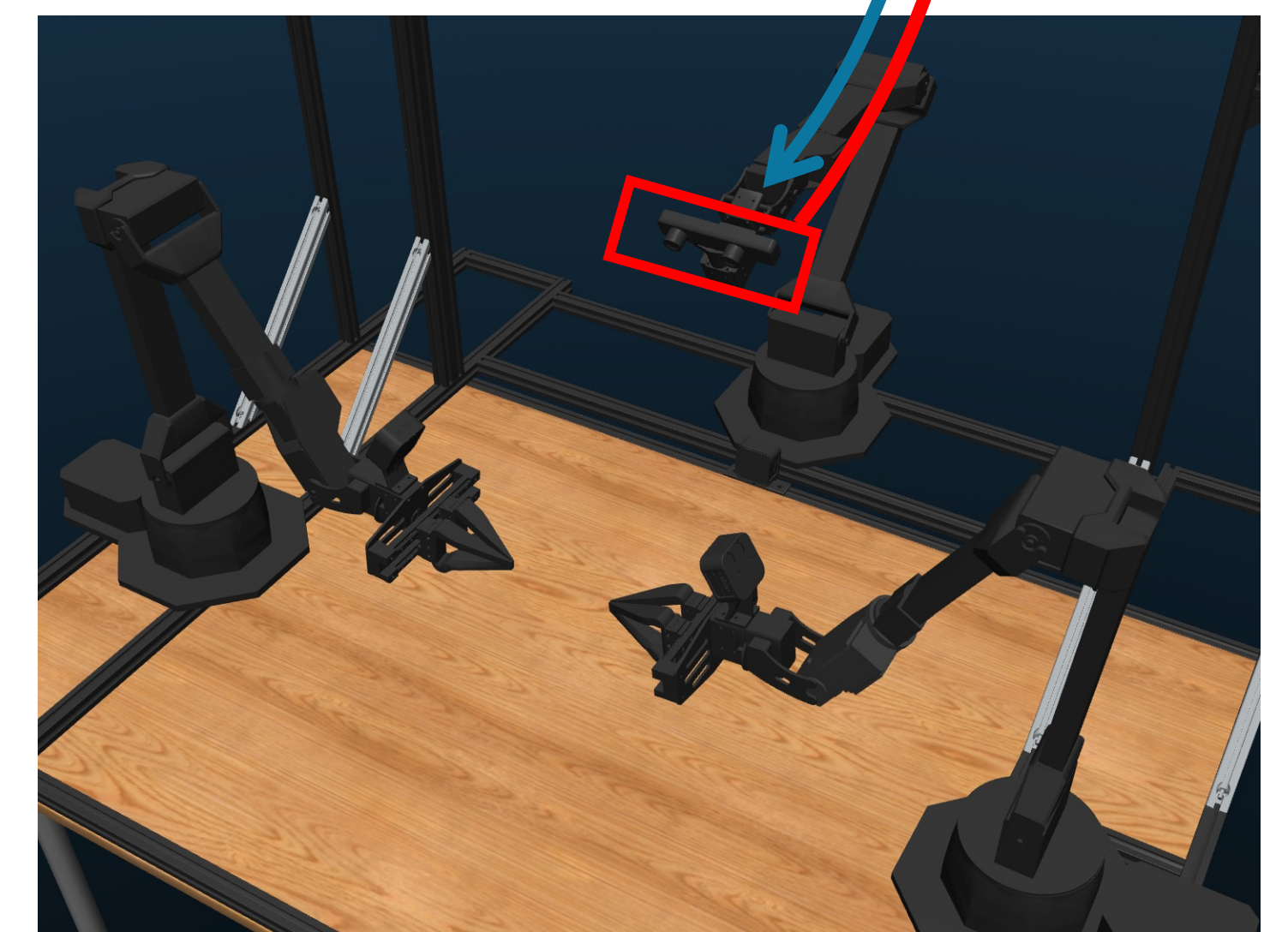
Is human-like **Active Vision** the missing piece in robotic perception? We developed **AV-ALOHA** to find out!



AV-ALOHA real-world setup



Human operators (Andrew & Ian)



AV-ALOHA simulation environment

AV-ALOHA

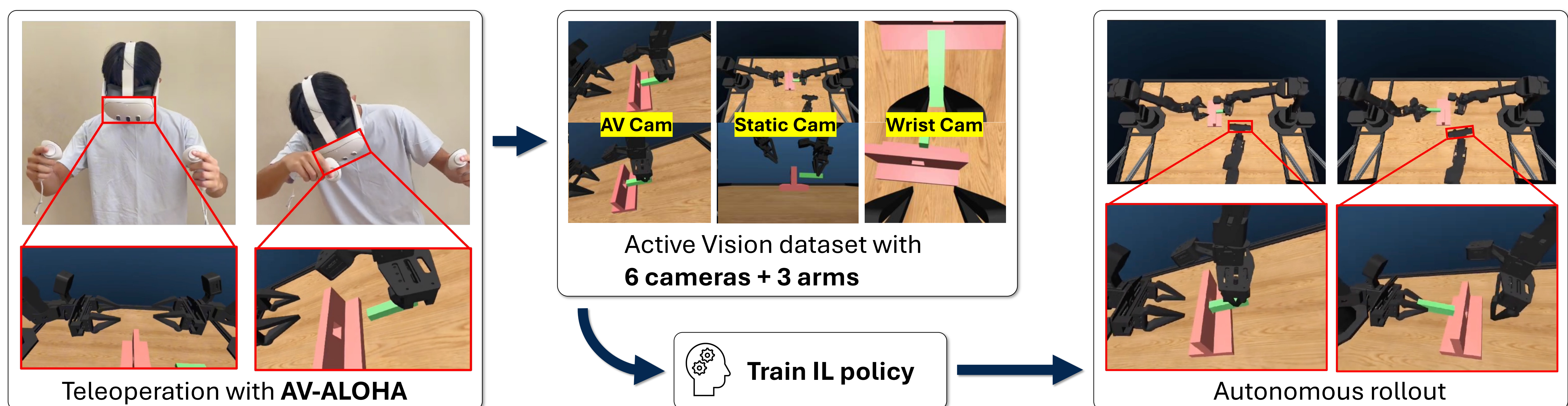
- A **bimanual robot system** with **7-DoF active vision (AV)** and **first-person VR control**
- Real & Sim platforms to explore learning **human-guided active vision**
- Open-source code, hardware, **simulation environment**, and datasets

WHY? Fixed-view cameras struggle with occlusions and limited visibility

HOW? Humans naturally move their heads and body to gain better perspectives

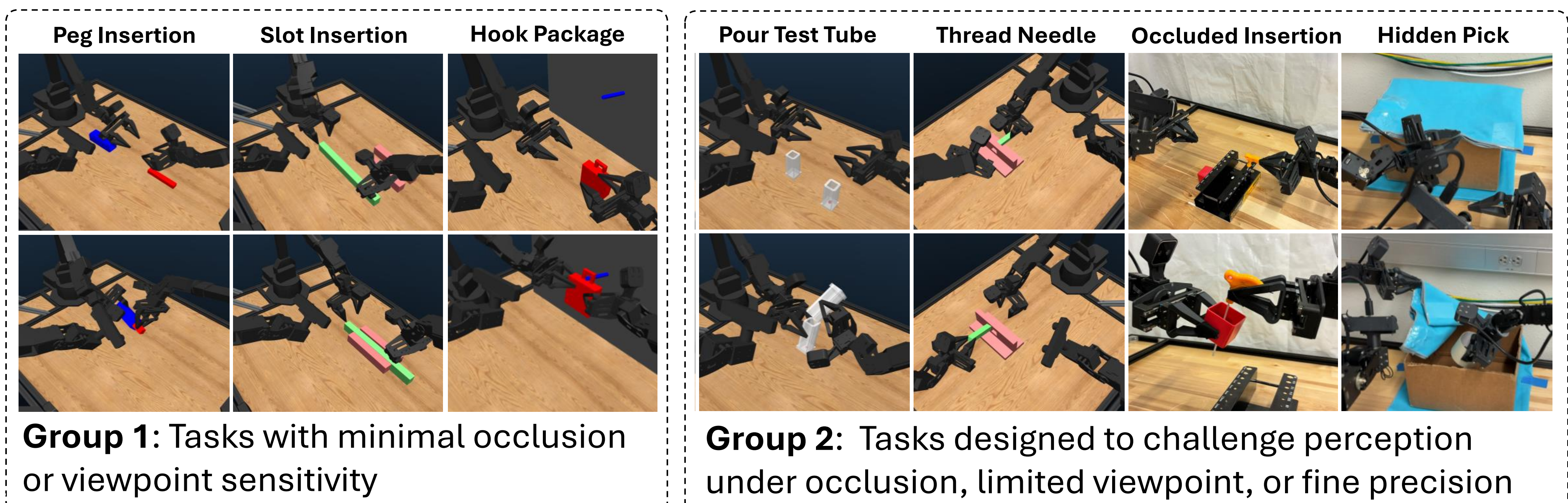
- **AV-ALOHA** can replicate this behavior — **actively control its viewpoint during demonstration and learning**

Imitation Learning Pipeline



Results

- We evaluate ACT success across seven tasks from two groups under different camera configurations
- Camera setups that include the **AV camera** outperform all others on every Group 2 task
- The **AV camera alone** achieves the highest success on two Group 2 tasks



★ **Active vision** enables perception strategies that adapt to shifting visual demands throughout the task



Detailed analysis of the results are in our paper

Camera Config	Group 1						Group 2							
	Peg Insertion		Slot Insertion		Hook Package		Pour Test Tube		Thread Needle		Occluded Insertion		Hidden Pick	
	Grasp	Insert	Grasp	Insert	Grasp	Hook	Grasp	Pour	Grasp	Thread	Grasp	Insert	Reveal	Pick
AV	74	42	88	50	100	22	66	14	98	52	60	20	95	55
AV + Static	84	46	100	62	100	34	50	10	98	26	20	0	90	45
AV + Wrist	82	34	96	44	100	22	70	14	92	52	95	30	95	60
AV + Static + Wrist	78	36	100	36	100	24	36	8	90	40	40	5	55	15
Static	84	48	98	66	100	44	44	8	88	30	85	20	100	40
Static + Wrist	88	40	100	78	100	30	46	6	38	22	100	15	100	35
Wrist	84	42	98	44	92	8	44	10	94	44	60	15	95	35