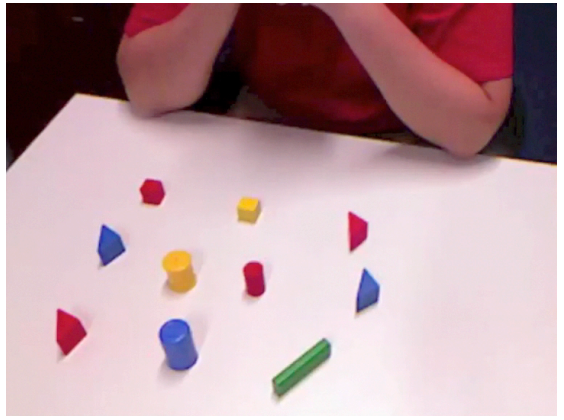
EAGER: Learning Language in Simulation for Real Robot Interaction

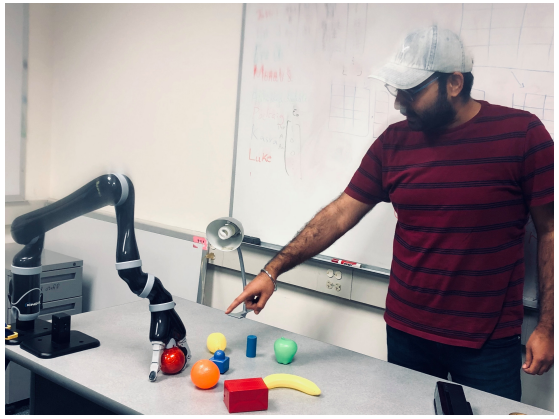
Cynthia Matuszek, Francis Ferraro, Don Engel
cmat, ferraro, donengel @ umbc.edu
University of Maryland, Baltimore County
iral.cs.umbc.edu

NSF #1940931

Collaboration requires communication

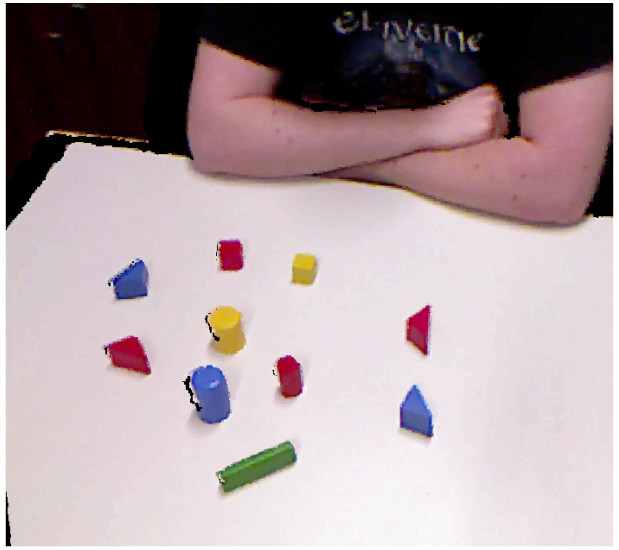
- Natural language for HRI
  - Natural, intuitive, and adaptable
- Teach, direct, and customize environment and teammates
- Language in the world = grounded language
  - NLP ≠ vision ≠ robotics
  - Grounded percepts and actions

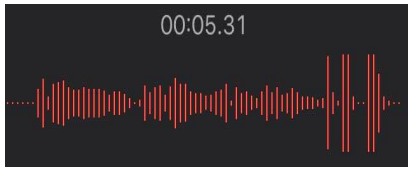




Jointly modeling correspondences

"Three cylinders; a red cylinder, a blue cylinder and a yellow cylinder."

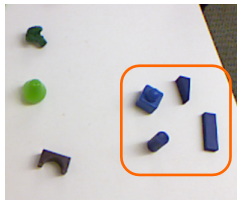




"These are the blue ones"

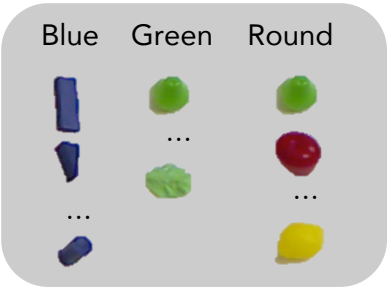
Language model

$\lambda x.newclass_0\text{-}blue(x)$



Grounded Query

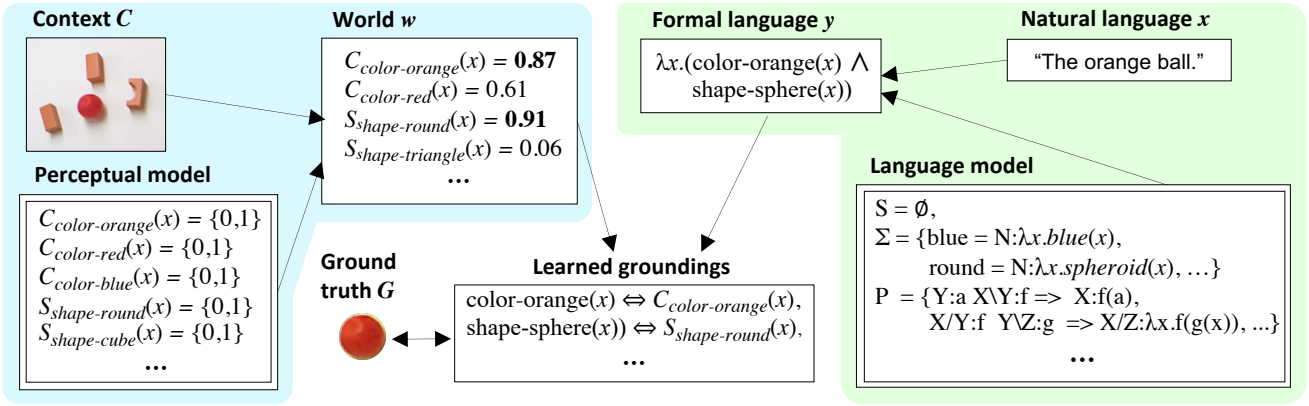
Perceptual world model

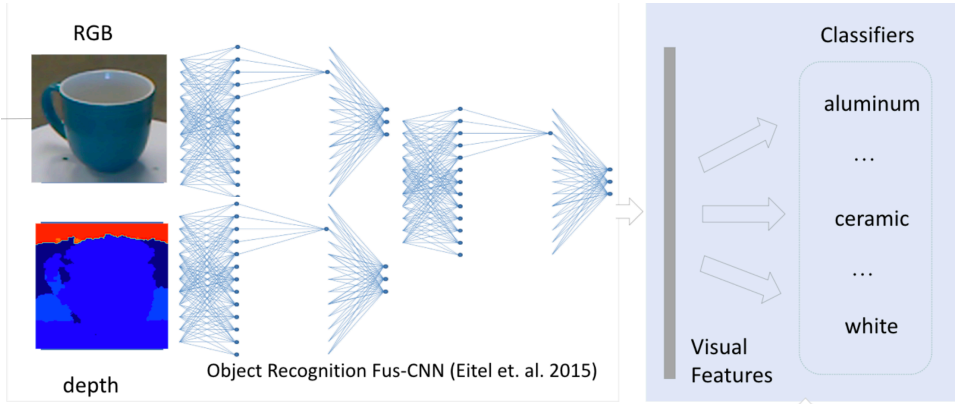


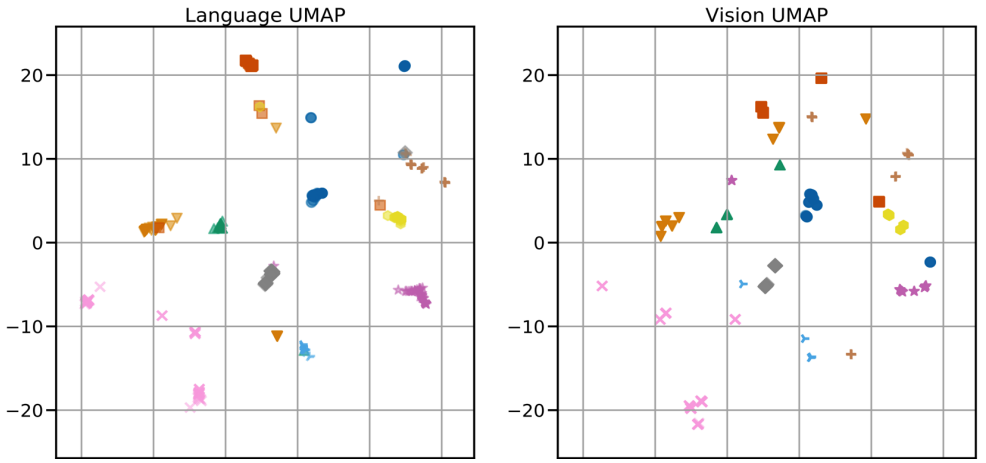
$$\underbrace{P(G,z,w \mid x,O)}_{\text{Joint Probability}} = \underbrace{P(z \mid x)}_{\text{Parsing Model}} \prod_{o \in O} \prod_{c \in C} \underbrace{P(w_{o,c} \mid o)}_{\text{Perceptual Model}} \underbrace{P(G \mid z,w)}_{\text{Grounding Query}}$$

The data problem(s)

- Formalism is model-agnostic
- Applicable to arbitrary modalities
  - As long as you have language
- But that makes it data hungry

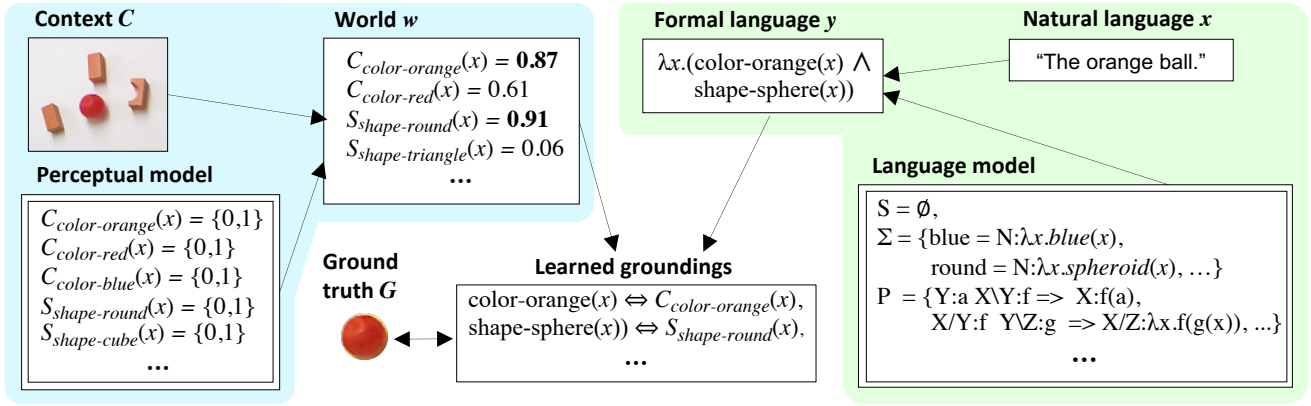


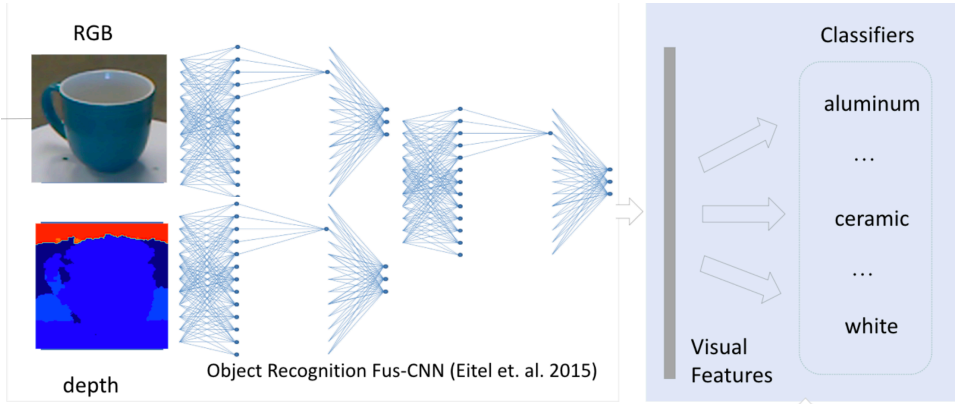


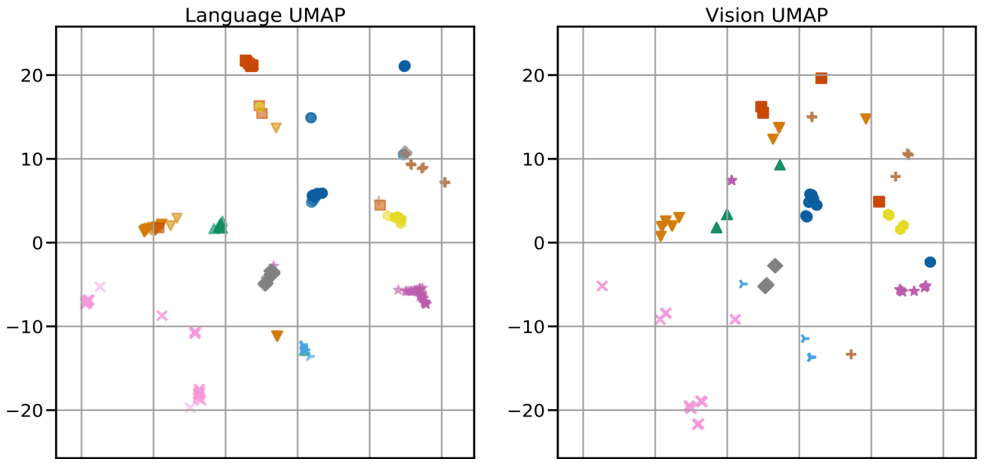


The data problem(s)

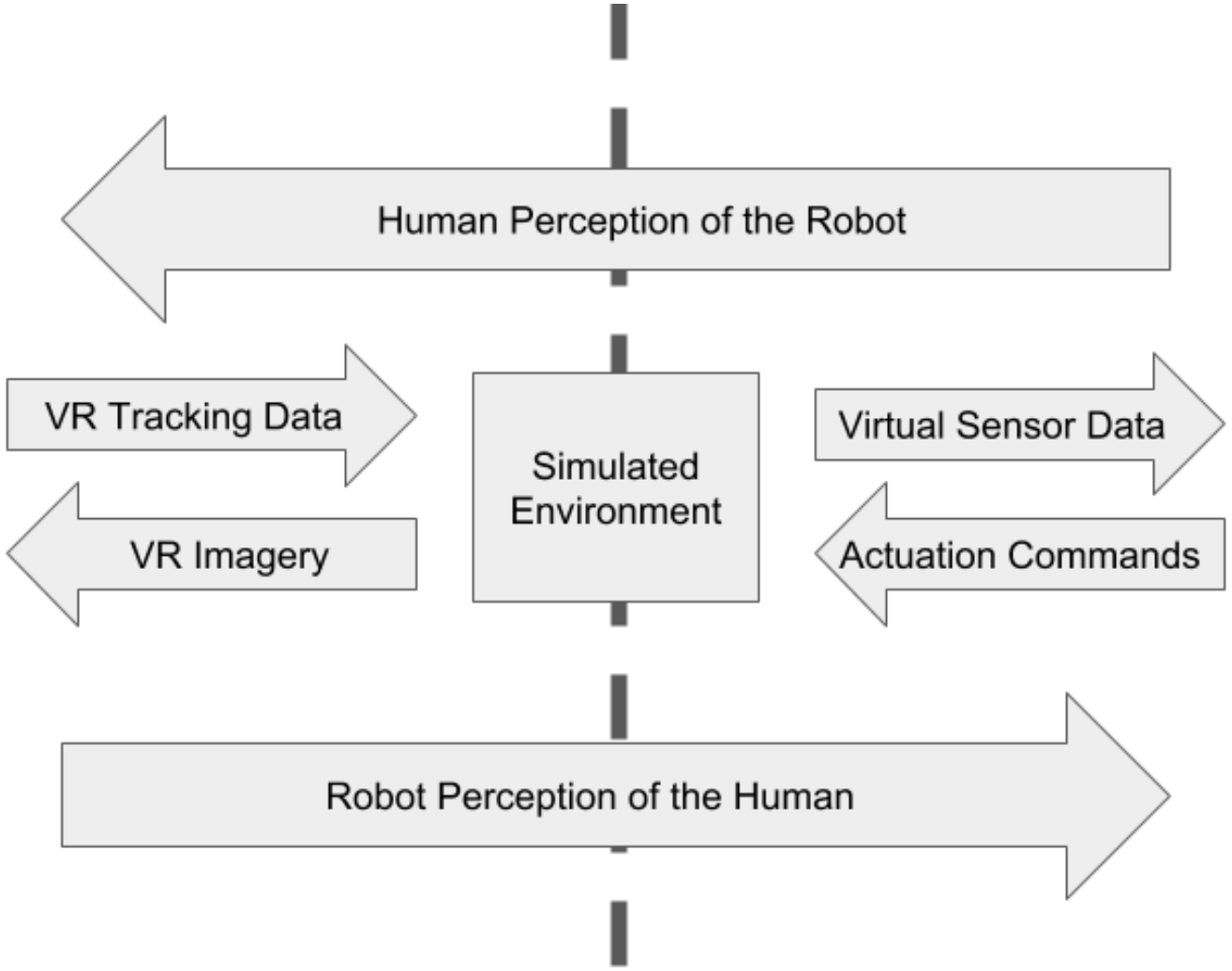
- Research platforms in large-*n* trials
- Collecting data from different settings
  - Getting robots & participants there, getting permission to be there, IRB, …
- Anything involving people







Sim2Real HRI, from the robot side



Training, simulating, testing

Instructing partner

Collect participant speech, gesture, and head posture

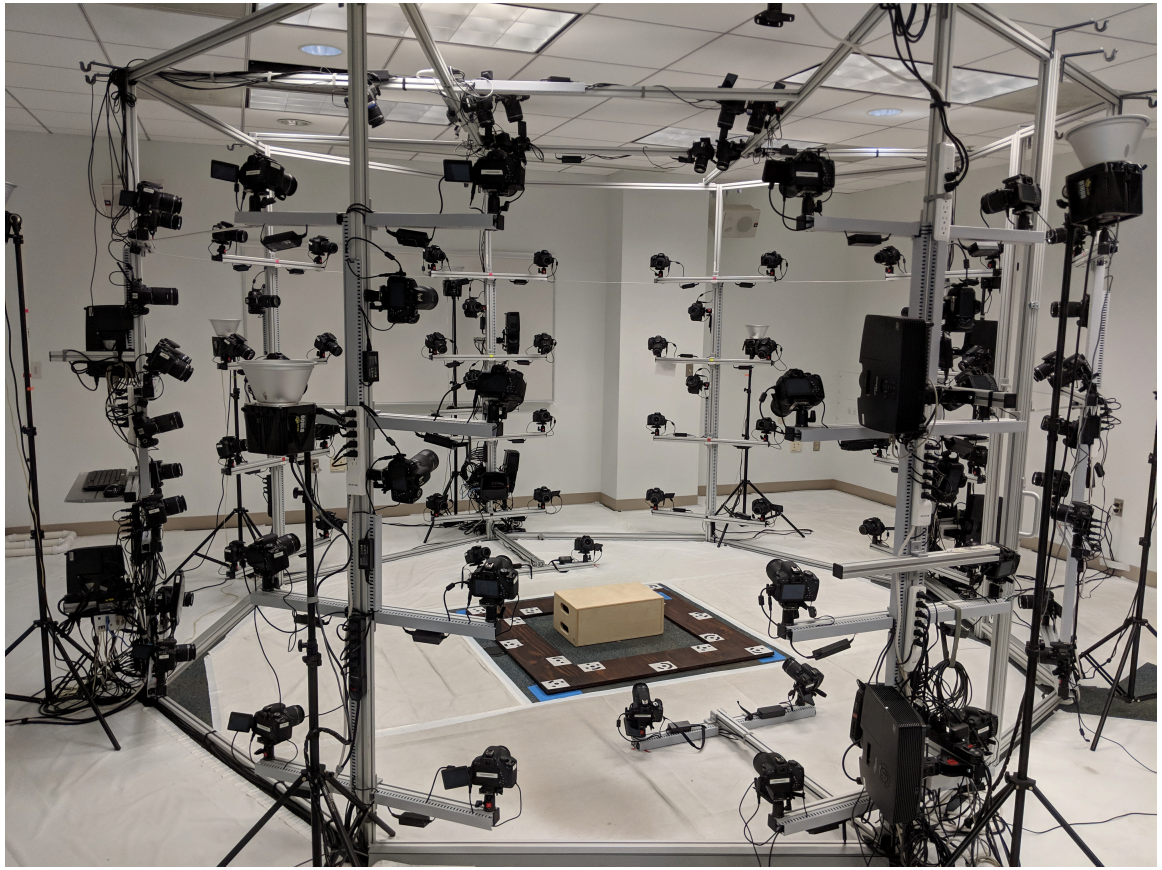
Robot's perceptual inputs

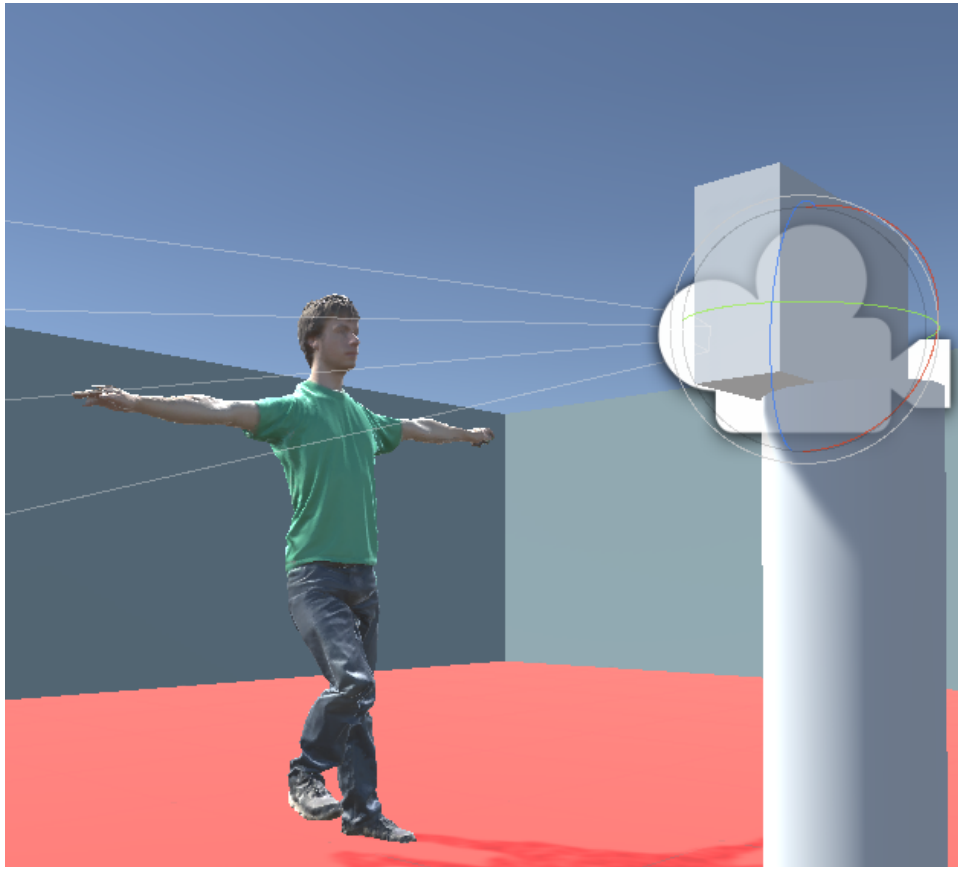
Saved scenario can be replayed with different sensors, noise model, …

Real-world transer

Perform tasks using domain adaptation for additional learning

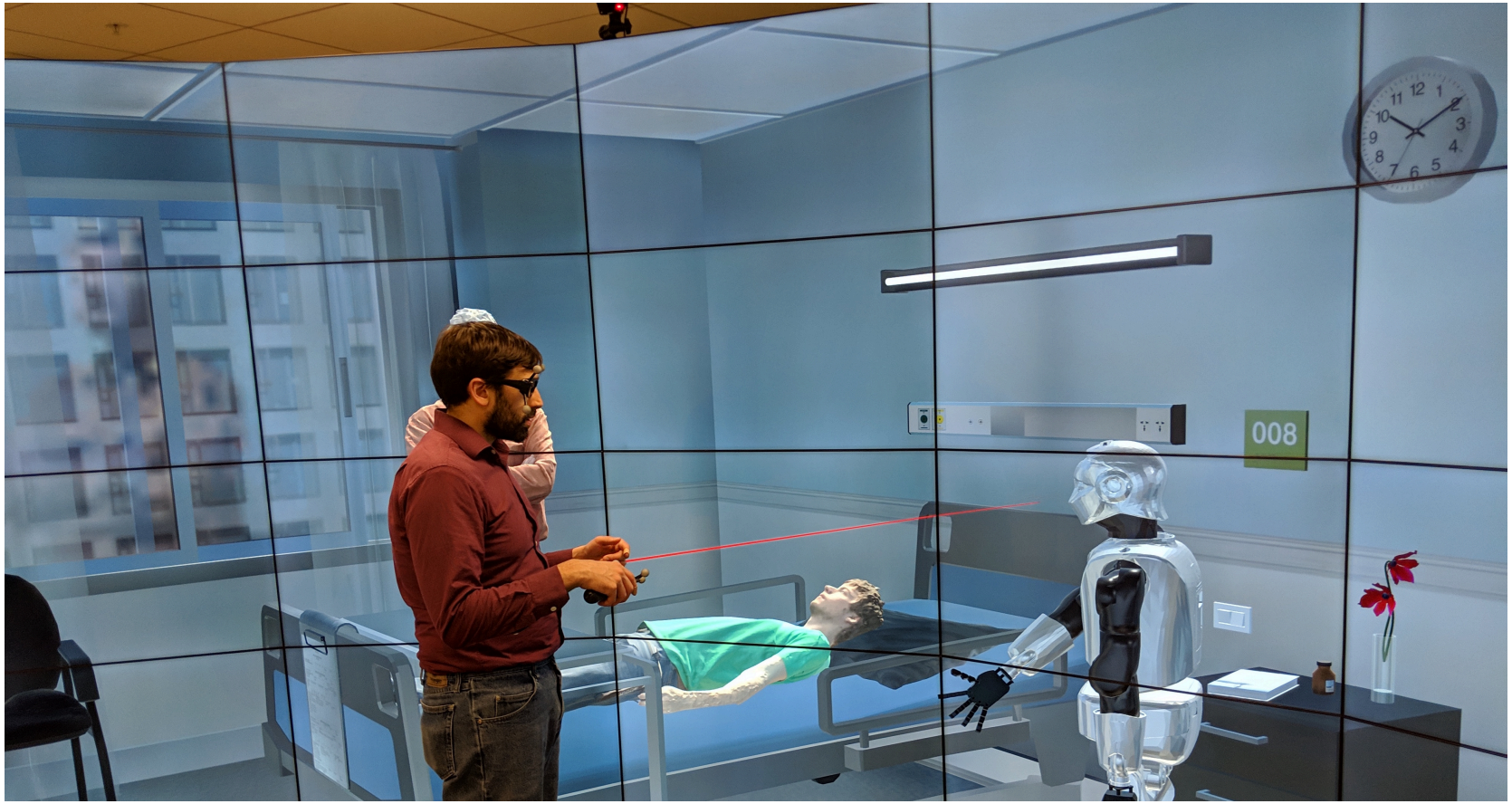
Photogrammetry: 3D models





NSF MRI #142820, PI: Marc Olano

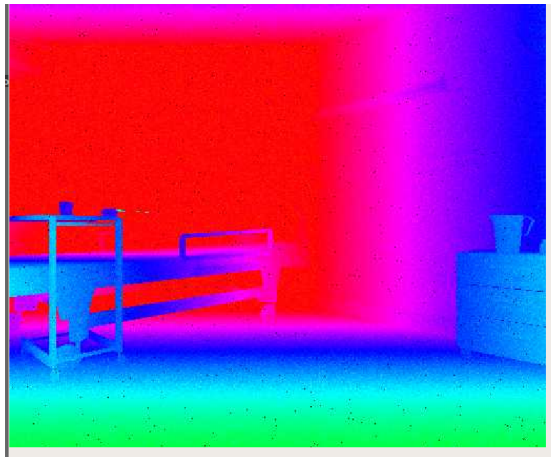
π^2: Participant interaction space

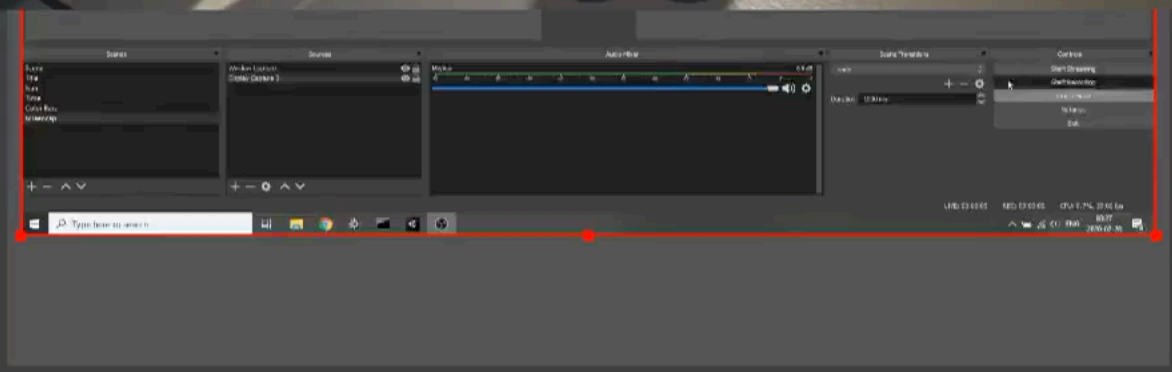


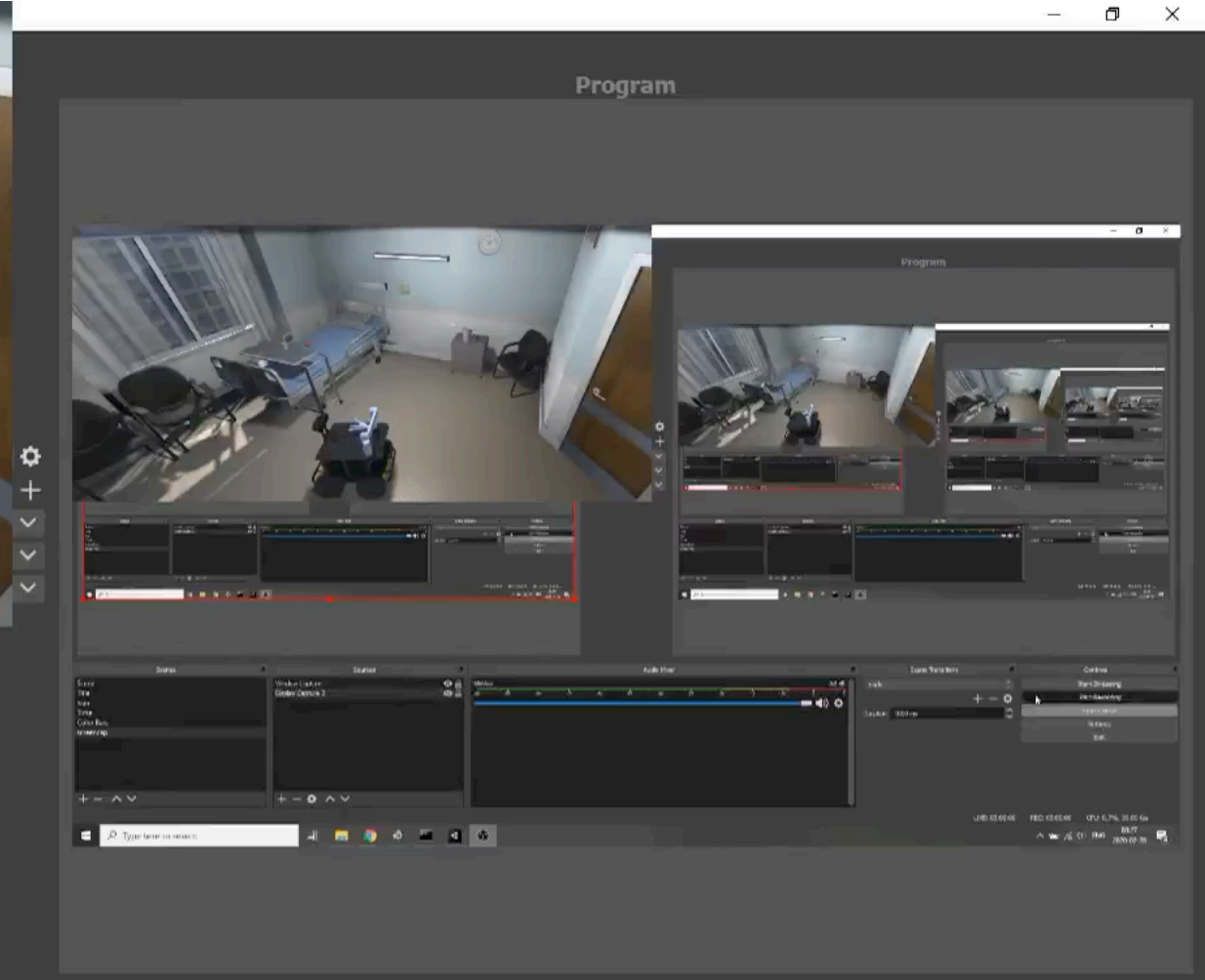
NSF MRI #1531491, PI: Don Engel

Scenario development and testing

- Fast experiment prototyping
- High-resolution data collection







Scenes

Scene
Title
Ivan
Time
Color Bars
screencap

Sources

Window Capture
Display Capture 2

Audio Mixer

Mic/Aux 0.0 dB

Scene Transitions

Fade
Duration 1000 ms

Controls

Start Streaming
Start Recording
Studio Mode
Settings
Exit

LIVE: 00:00:00 REC: 00:00:00 CPU: 0.7%, 30.00 fps

Type here to search

ENG 00:27 2020-02-28