

# RGB-D: Advanced Reasoning with Depth Cameras



Hema Koppula  
Cornell University



John Leonard  
MIT



Ashutosh Saxena  
Cornell University

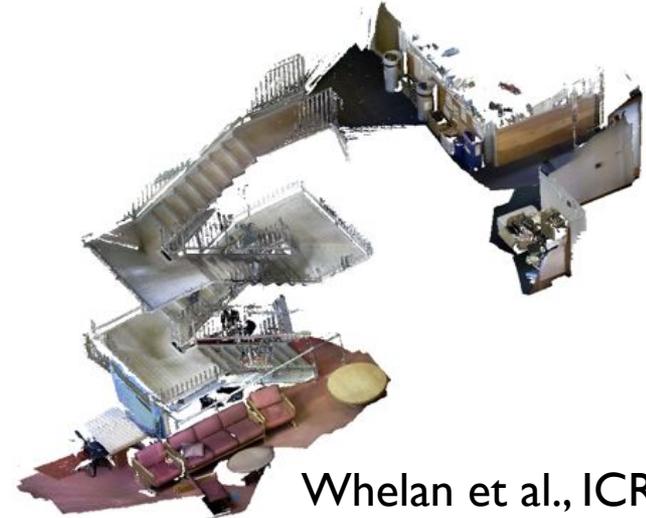
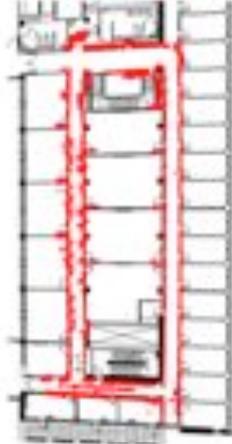


Jianxiong Xiao  
Princeton University

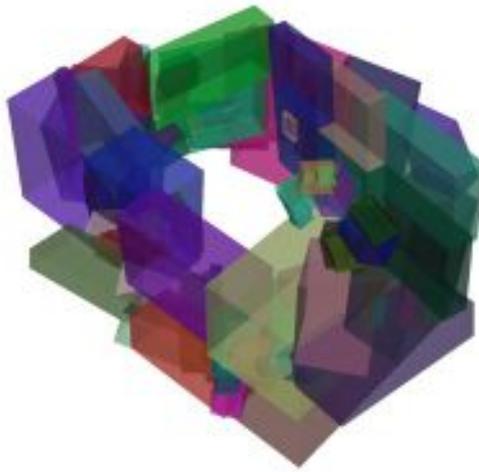
# RGB-D Sensors



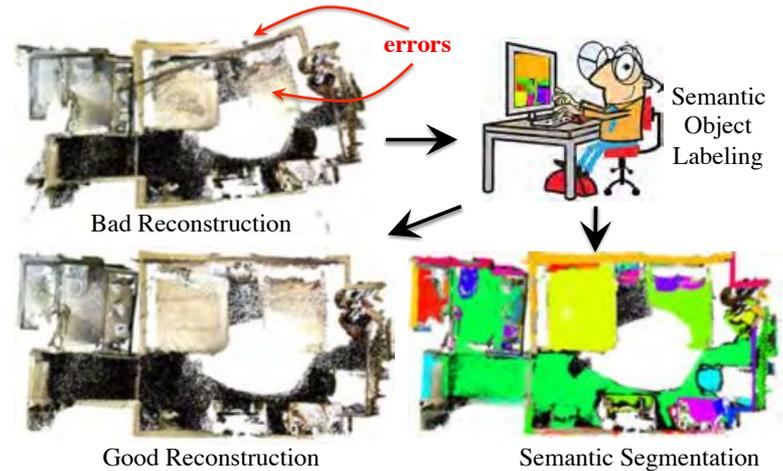
# 3D Reconstruction



Whelan et al., ICRA '13



Henry et al., 3DV '13, RSS-RGBD '13



Good Reconstruction

Semantic Segmentation

Xiao et al., ICCV '13

# Object Detection and Scene Labeling

---



Lai & Fox et al. ICRA '11, '12, '14



Koppula & Saxena et al. NIPS '11, IJRR '12

# Semantic Labeling with 3D Context

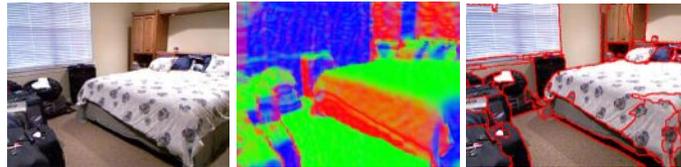
---

- ▶ Goal: Detect what objects are present
- ▶ Properties
  - ▶ Image + 3D shape
  - ▶ Object-object relations
  - ▶ Object part relations
- ▶ Principled learning algorithm
- ▶ Tested on a large dataset
  - ▶ 550 views from 52 homes and offices (*Cornell RGBD Dataset*)
  - ▶ Tested on multiple robots
  - ▶ Code and data available



**Our algorithm running on  
POLAR Robot to find a  
keyboard**

# Scene Understanding and Segmentation



Input RGB

Aligned Normals

Segmentation

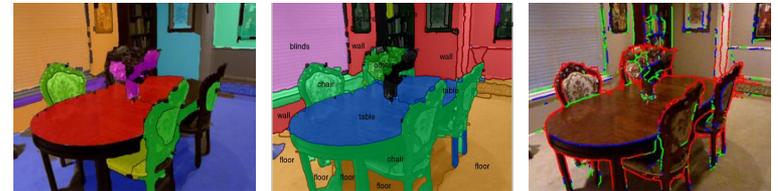


Input Depth

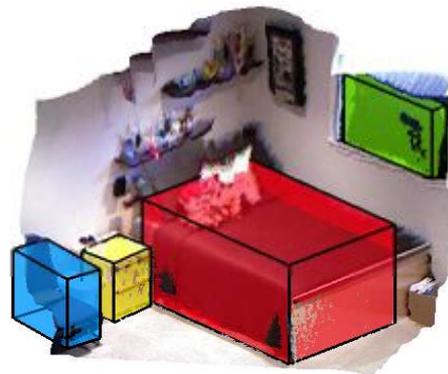
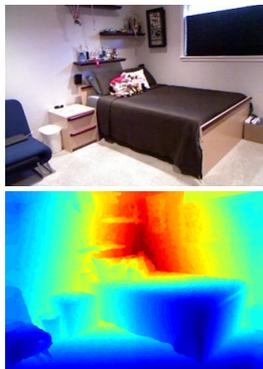
3D Planes

Support Relations

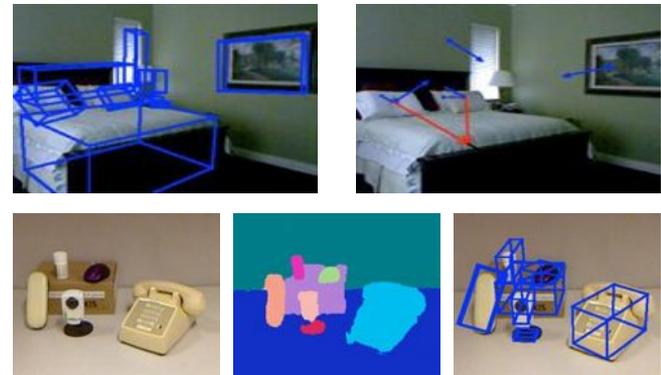
Silberman, Hoiem, Kohli & Fergus, ECCV'12



Gupta, Arbelaez & Malik, CVPR'13



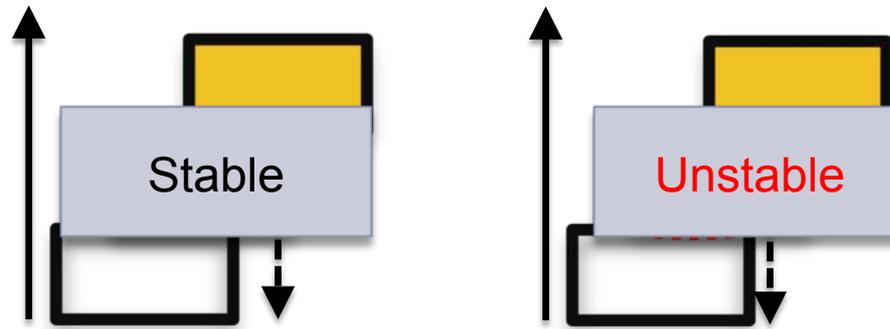
Lin, Fidler & Urtasun, ICCV'13



Jia, Gallagher, Saxena & Chen, CVPR'13

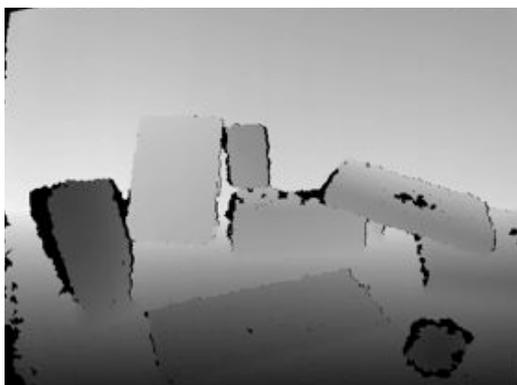
# Intuitive Physics: Stability

---



# Reasoning with Blocks, Support, & Stability

---

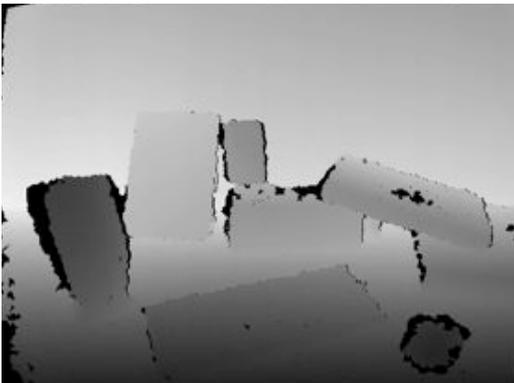


Input: RGB-D



Segmentation

# Reasoning with Blocks, Support, & Stability



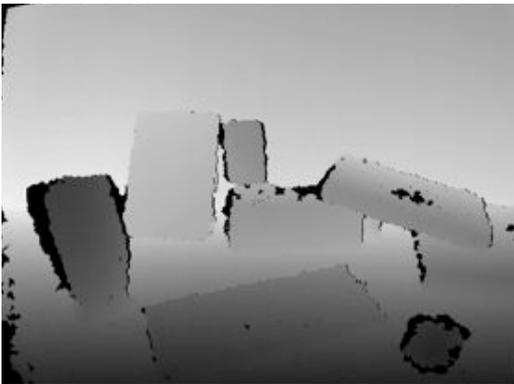
Input: RGB-D



Blocks, Support, and Stability

# Reasoning with Blocks, Support, & Stability

---

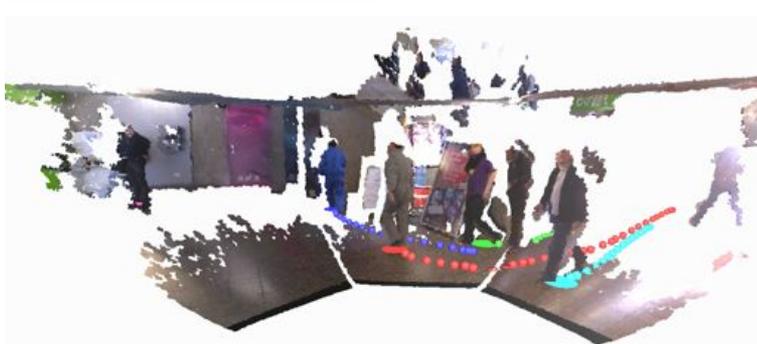
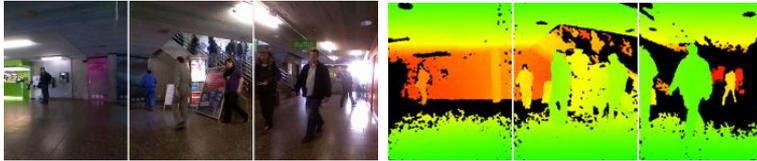


Input: RGB-D

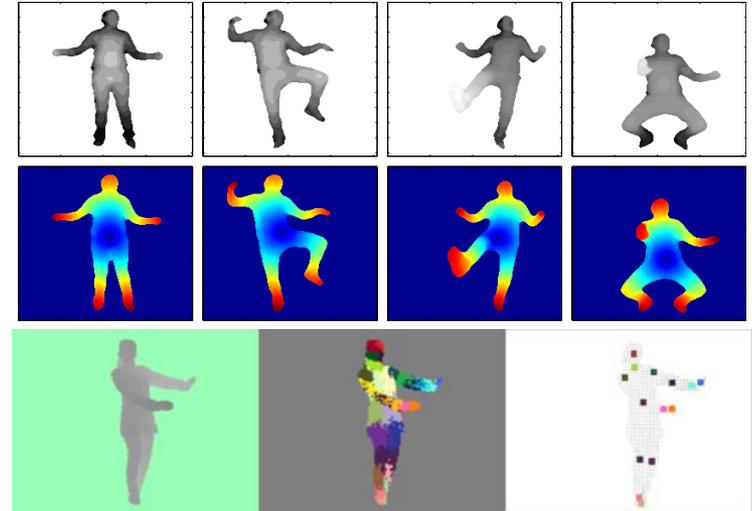


Final 3D representation

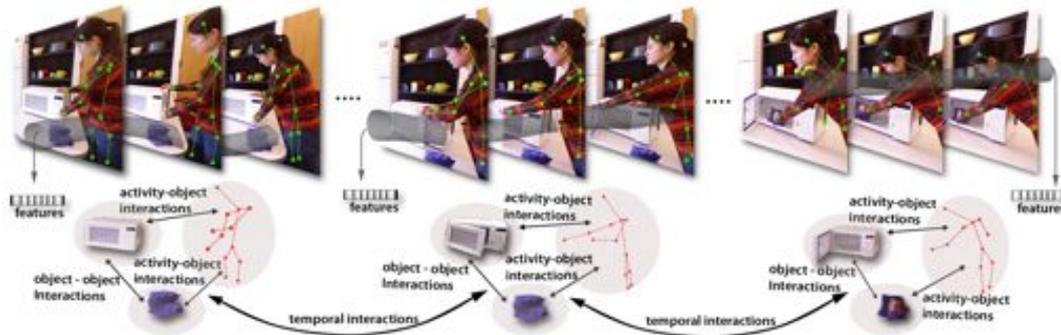
# Understanding Humans



Luber et al. IROS '11  
Munaro et al. IROS '12



Schwarz et al. FG '11  
Shotton et al. PAMI '12



Koppula, Gupta & Saxena, IJRR '13

# Human Activity Detection



Robot observes subject

Tested on CAD -120 dataset

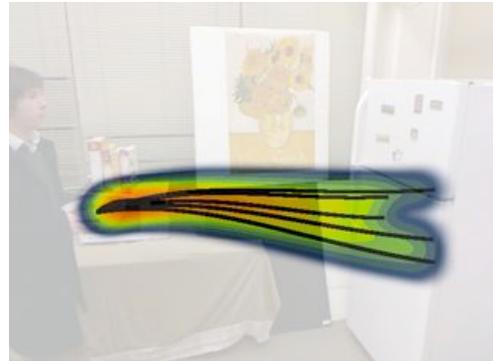
(Koppula, Gupta & Saxena, IJRR '13)

▶ *'having a meal', 'taking medicine', 'microwaving food', etc.*

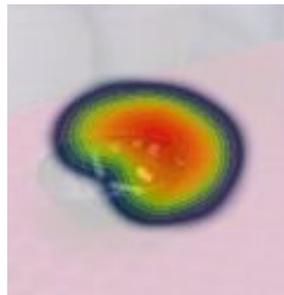
Hema Koppula

# Modeling humans and objects

## ▶ Human Intentions



## ▶ Object Interactions



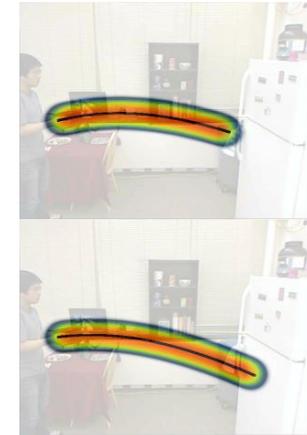
Bowl: pour-to Pitcher: pourable



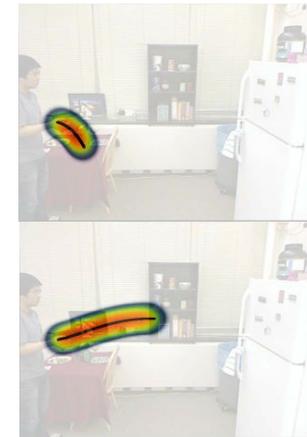
# Modeling Human Intentions



{*reachable*}

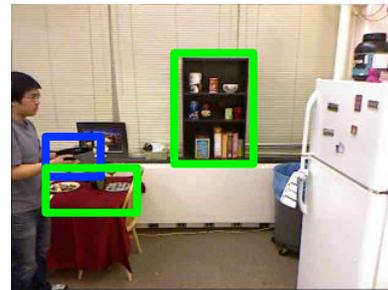
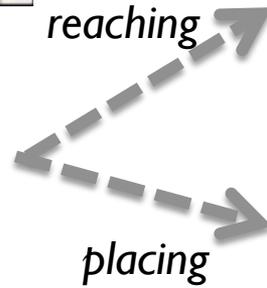
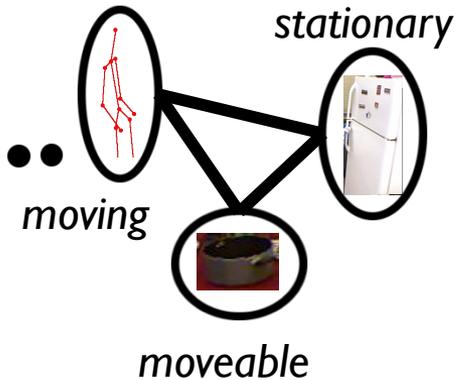


⋮



⋮

Top Anticipations



{*placeable*, *reachable*}

# Modeling Object Affordances

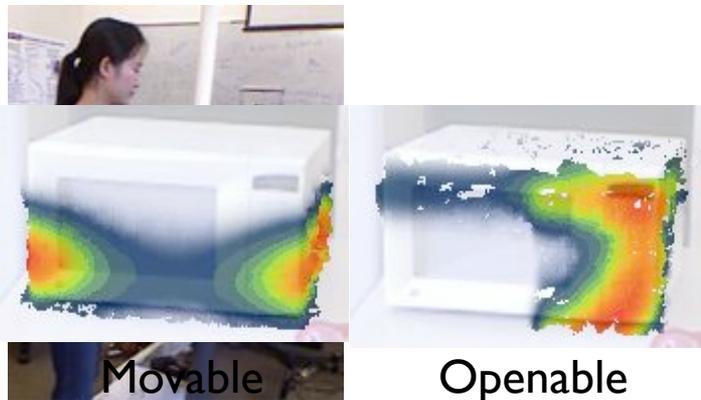
“All *action possibilities* latent in the environment.”

- J.J. Gibson (1977)

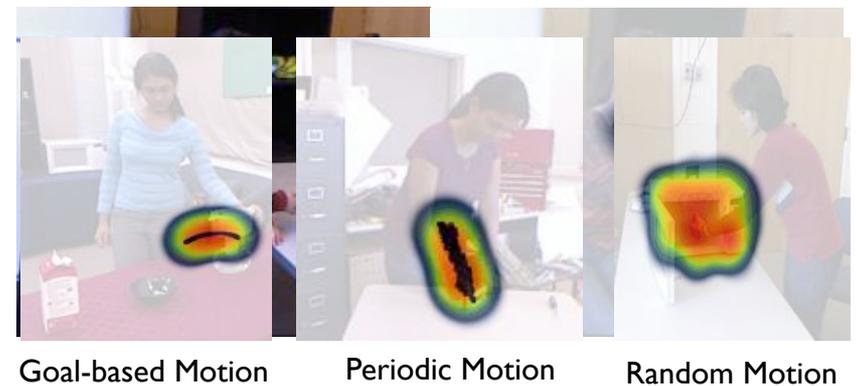
## Semantic Labels



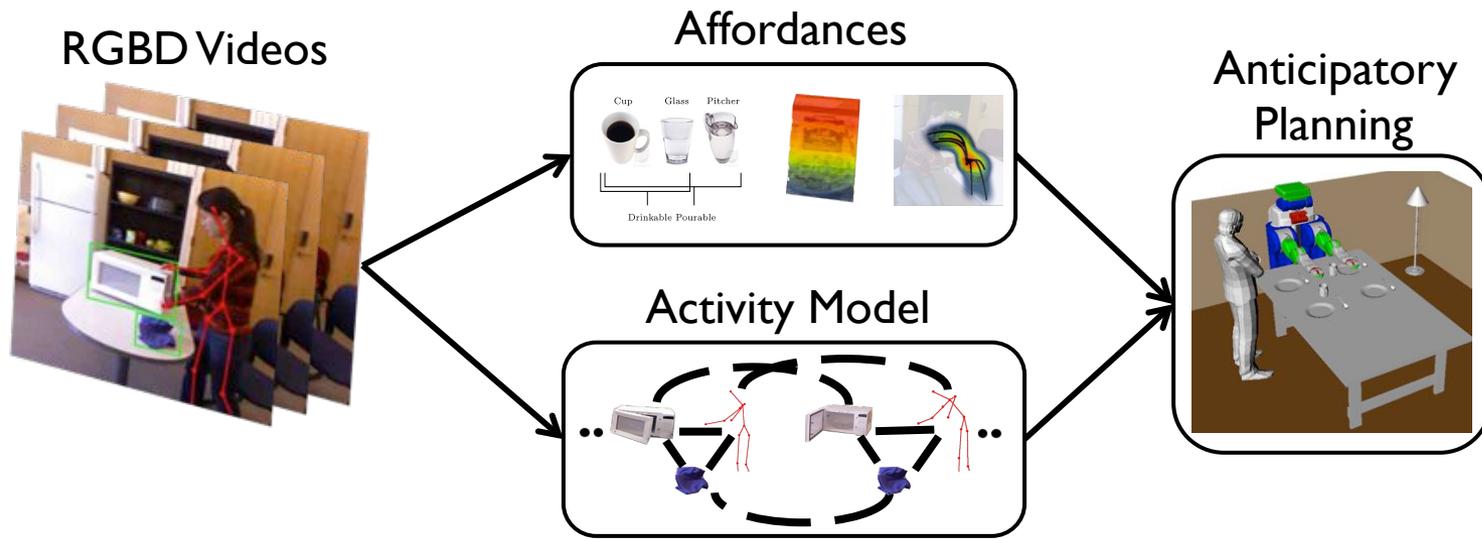
## Spatial Interactions



## Temporal Trajectories



# From Perception to Planning

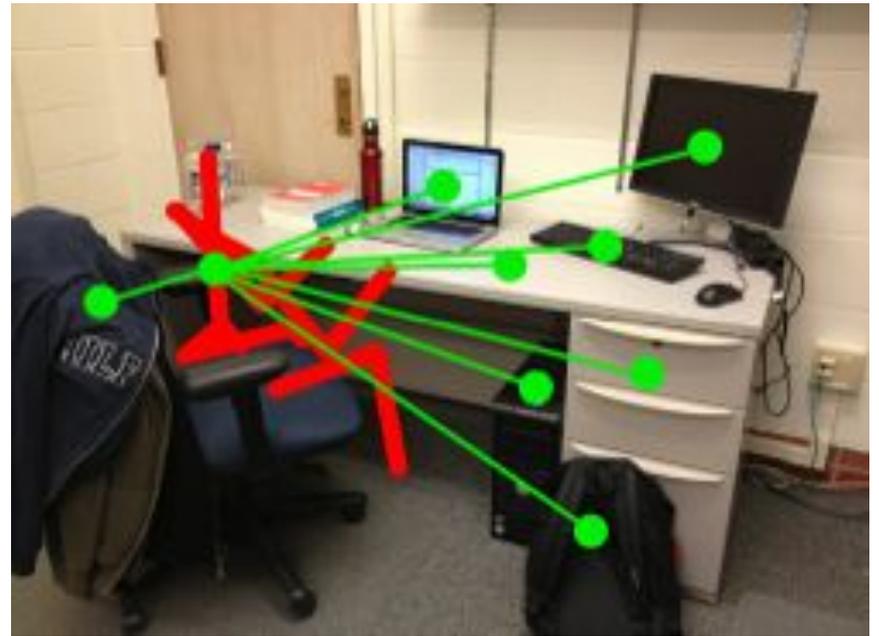
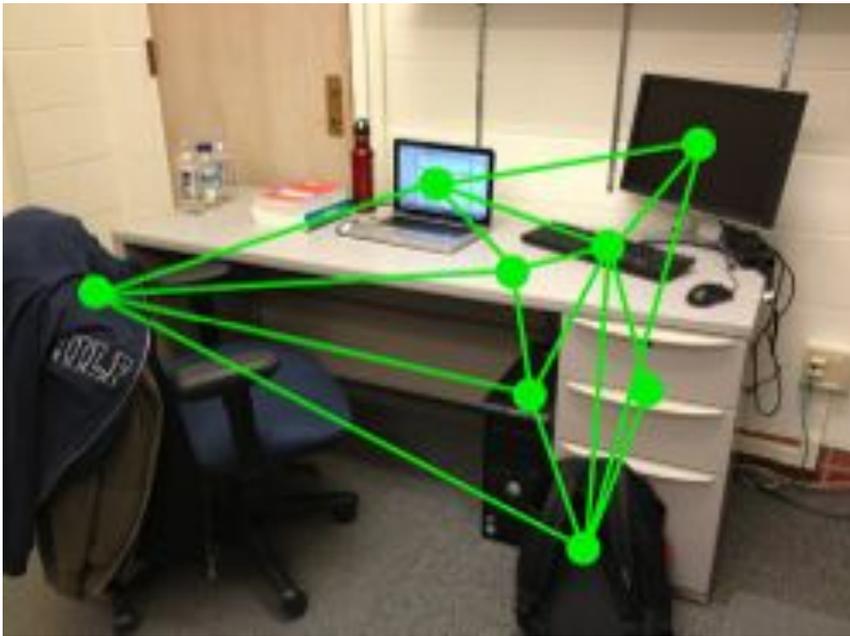


Human + Environment



# Application of Human Modeling to Object Detection

---



# Program

---

- ▶ 6 Invited Talks
- ▶ 9 Contributed Papers
- ▶ Panel Discussion ( 12:00 – 12:30 pm )
- ▶ Discussion ( 4:30 – 5:00 pm )
- ▶ Poster Session ( 5:00 – 6:00 pm)

# Invited Speakers

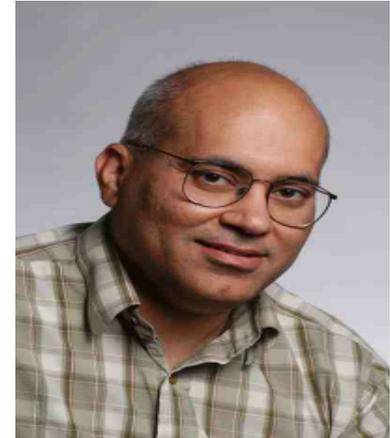
---



**Dieter Fox**



**Martial Hebert**



**Jitendra Malik**



**Derek Hoiem**



**Noah Snaveley**



**Raquel Urtasun**