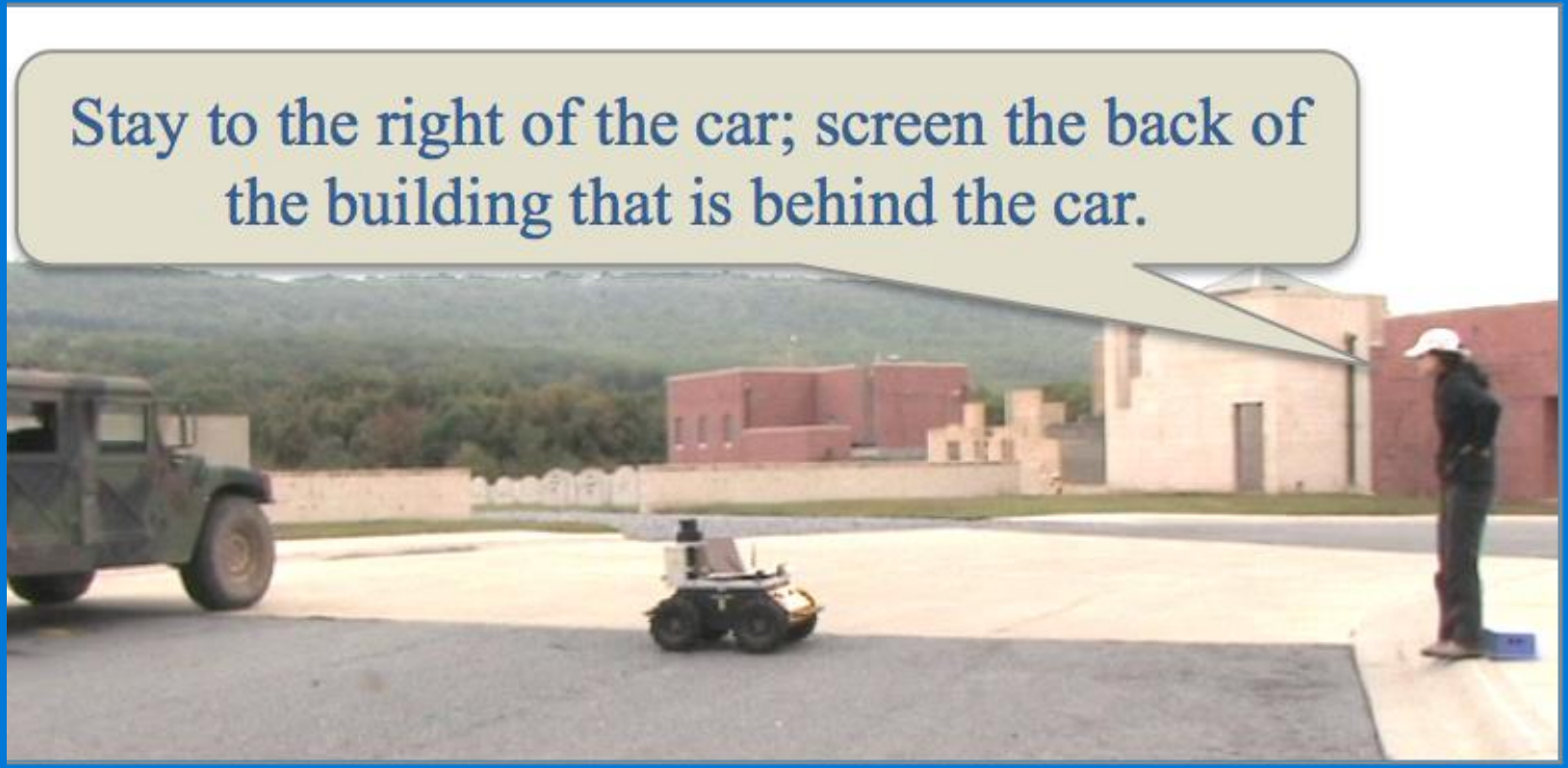
Computer vision for autonomy

Martial Hebert
The Robotics Institute
CMU

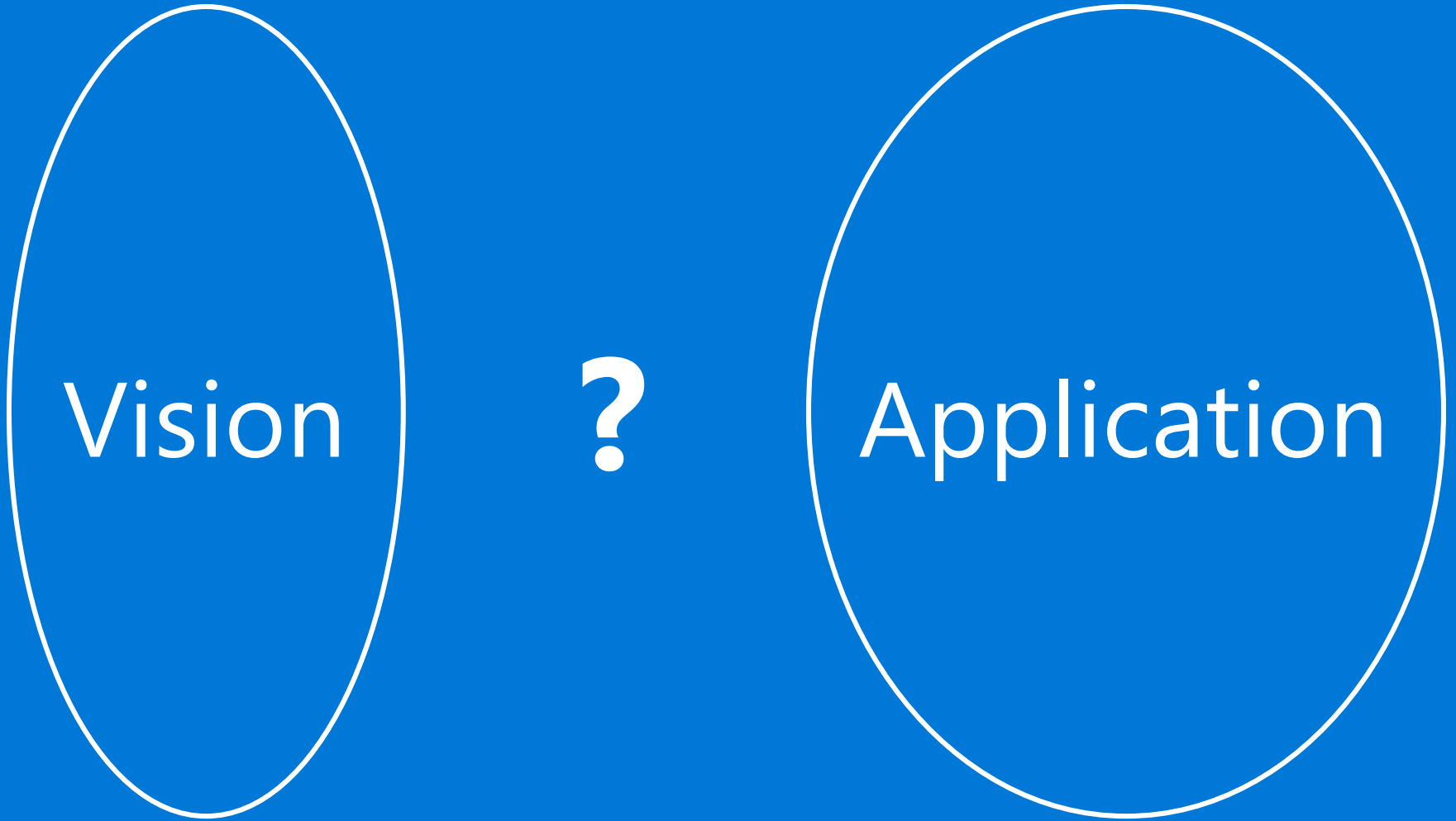
sky tree road sidewalk building car

column_pole pedestrian bicycle fence sign_symbol


Stay to the right of the car; screen the back of the building that is behind the car.

RCTA Semantic Navigation

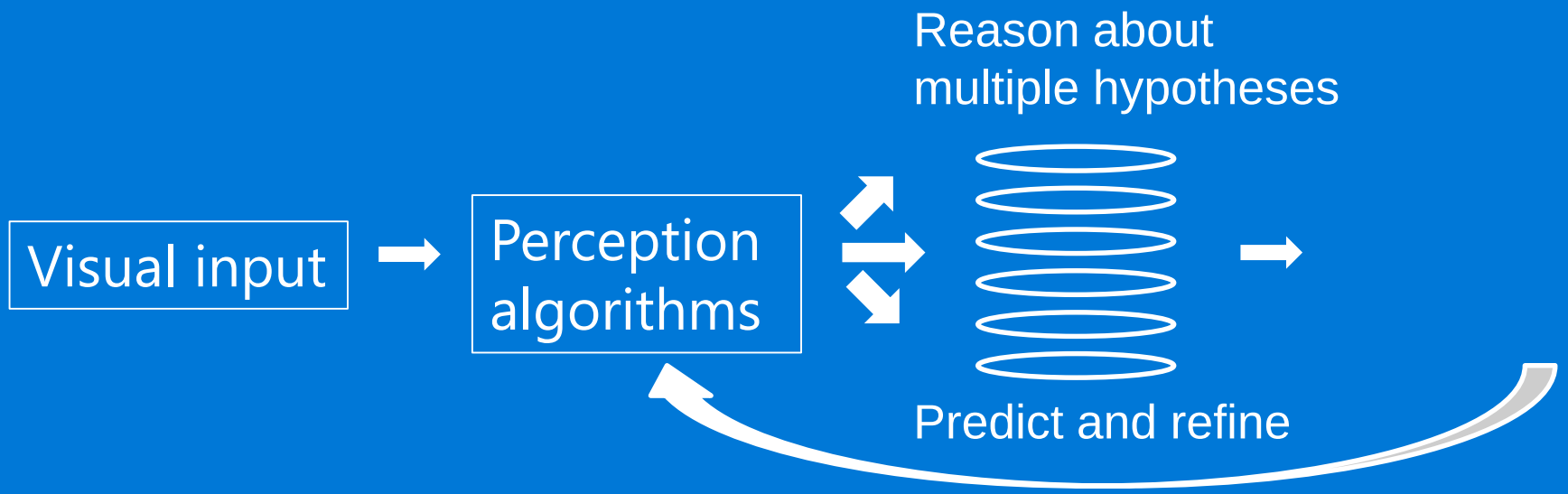
Jean Oh, Arne Suppé, Luis Navarro-Serment,
Felix Duvallet, Abdeslam Boularias,
Oscar Romero, Jerry Vinokurov,
Christian Lebiere, Martial Hebert, Anthony Stentz
Carnegie Mellon University
Robert Dean, Terence Keegan, Chip DiBerardino
General Dynamics Robotic Systems

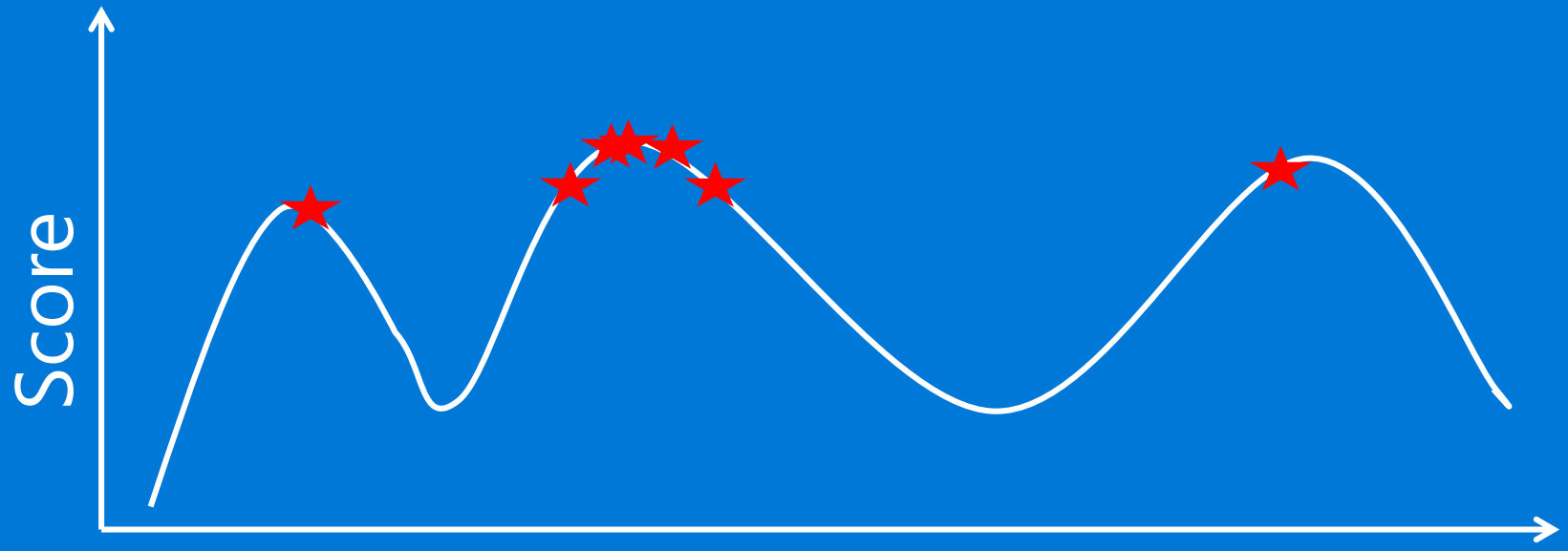

Vision
?
Application

Microsoft

Don’t commit early

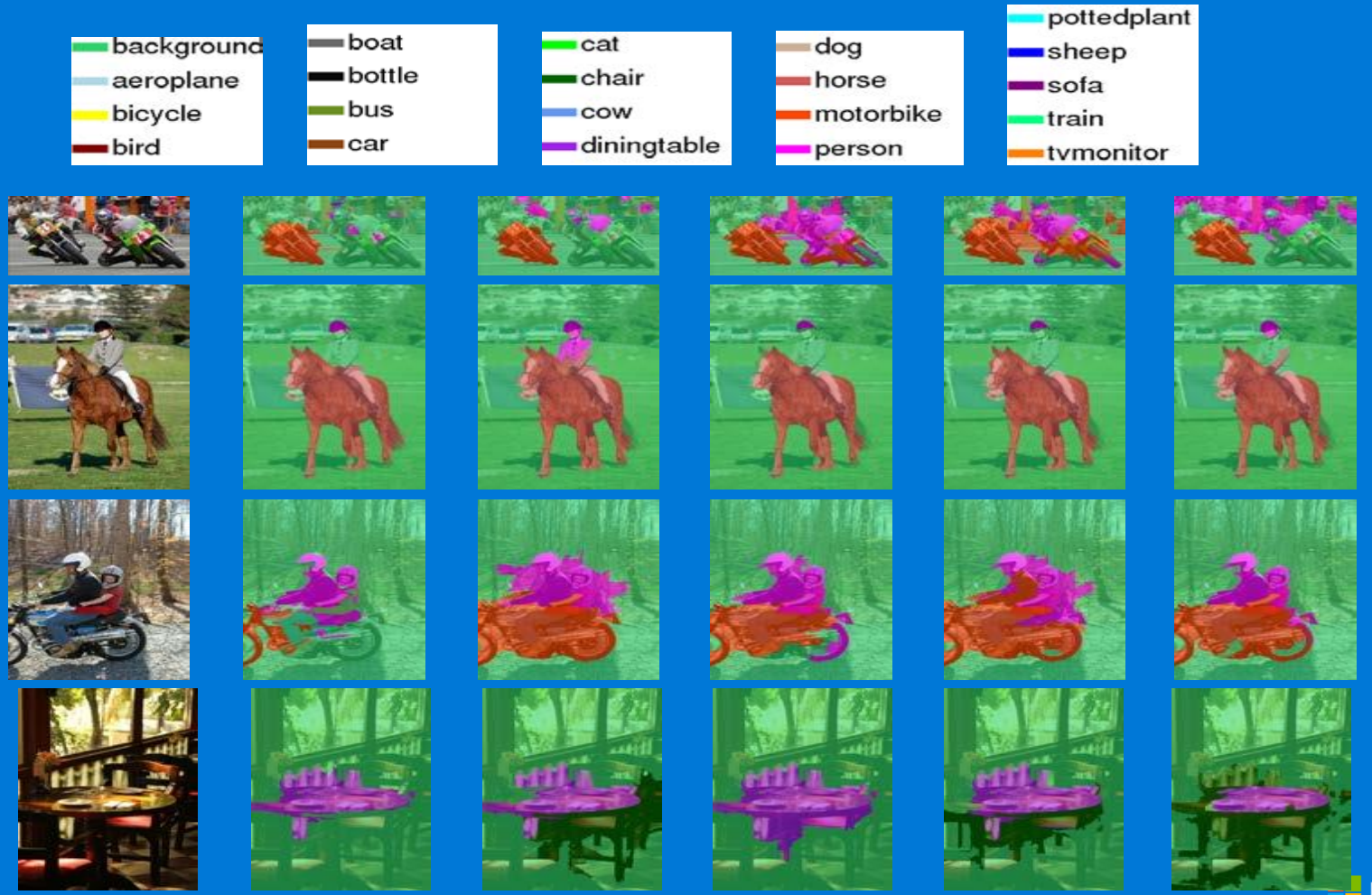

Visual input
Perception algorithms
Interpretation


Reason about
multiple hypotheses
Visual input
Perception
algorithms
Predict and refine



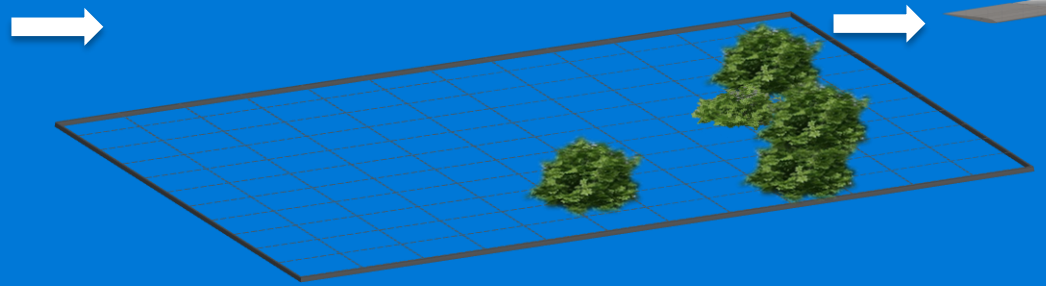
- Likely hypotheses
- Diverse hypotheses
- Complex, unobservable distribution

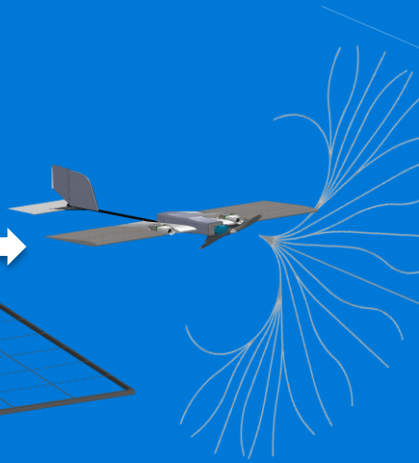
Microsoft



Microsoft

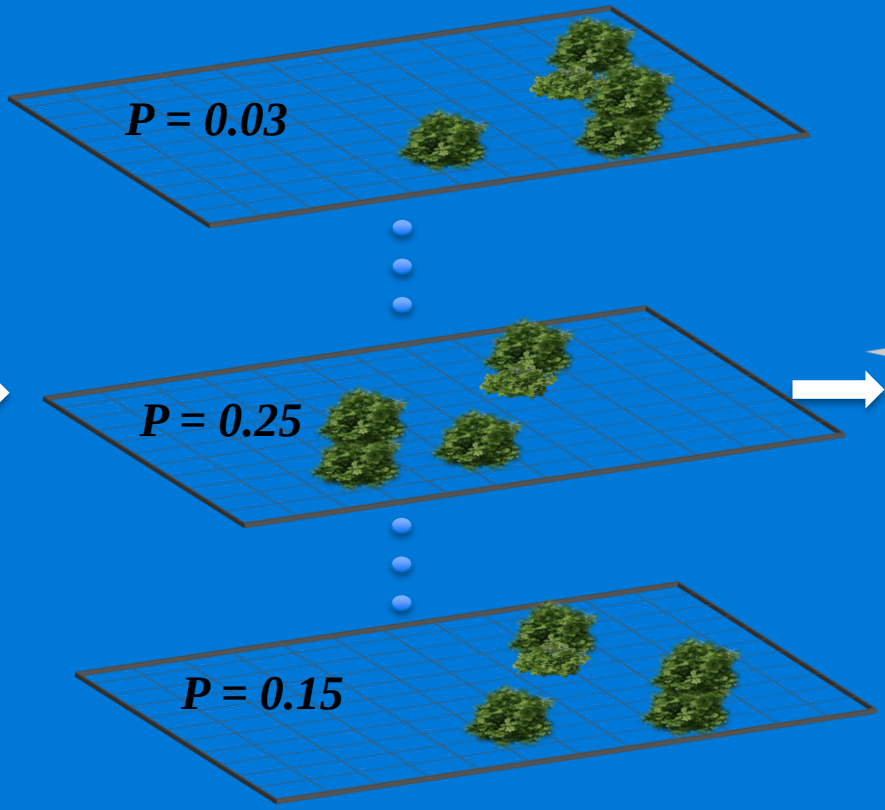
21 cl_second-order



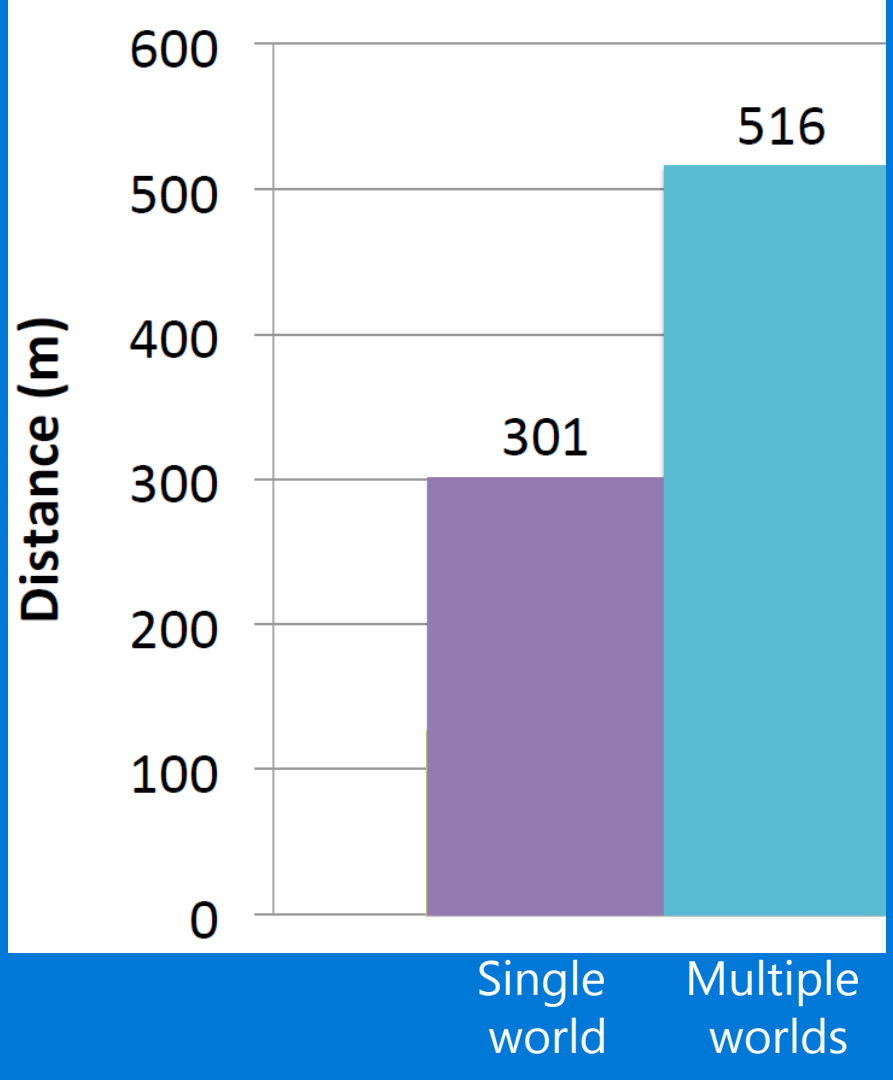


Debadeepta Dey et al. Vision and Learning for Deliberative Monocular Cluttered Flight. FSR 2015

Microsoft

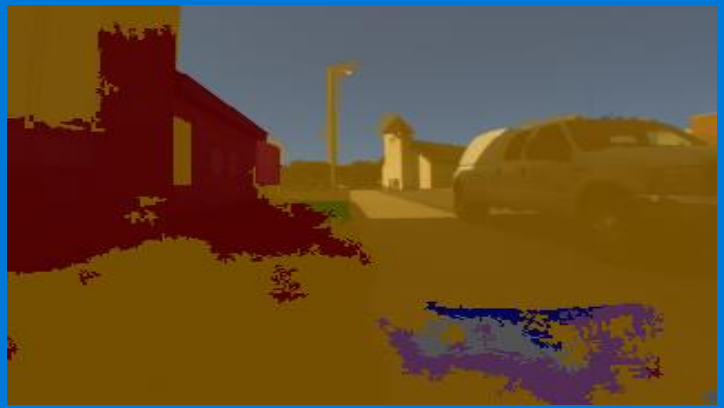


Debadeepta Dey et al. Vision and Learning for Deliberative Monocular Cluttered Flight. FSR 2015

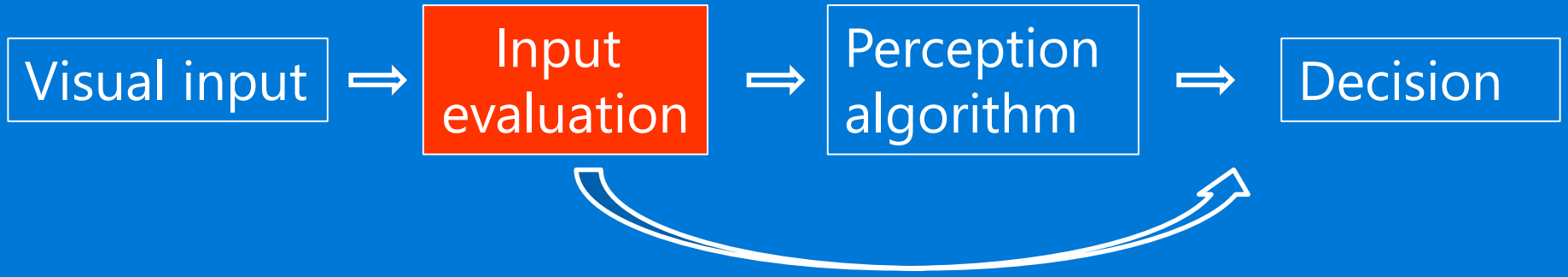

600
500
400
300
200
100
0
Distance (m)
516
301
Single world
Multiple worlds

Always give an answer?




Lennon et al. Performance Evaluation of a Semantic Perception Classifier. ARL-TR-6653

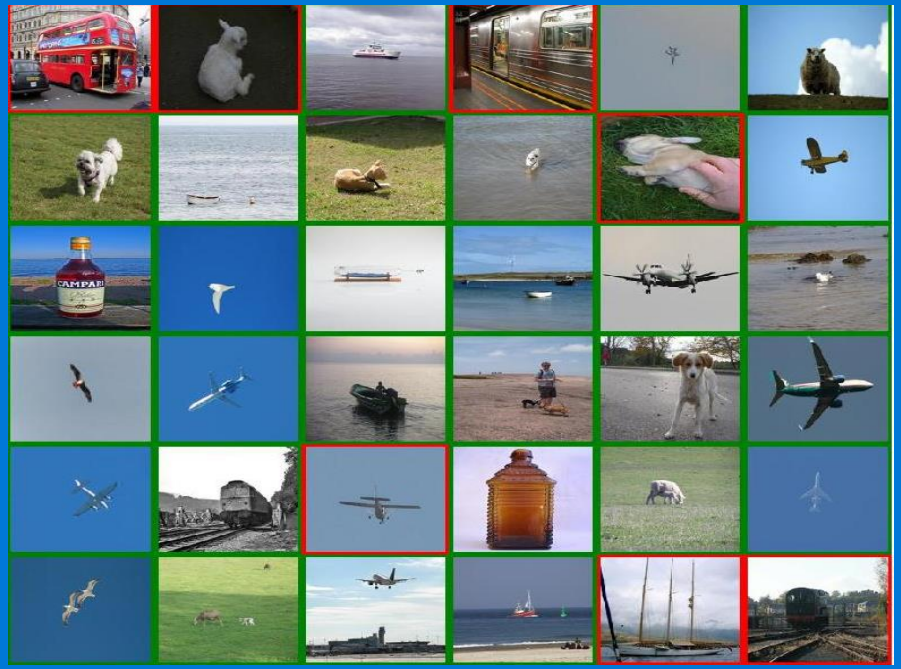


Input features → Accuracy

(e.g., RBF regressor,
SVM with prob. output)

With J. Wang, D. Parikh, A. Faradhi, CVPR 2014

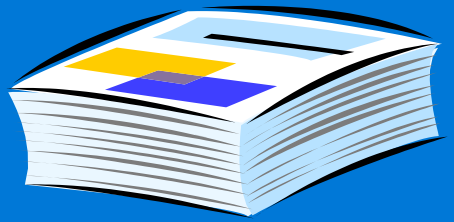
Microsoft



Microsoft

Use all available knowledge

Approximate locations/maps

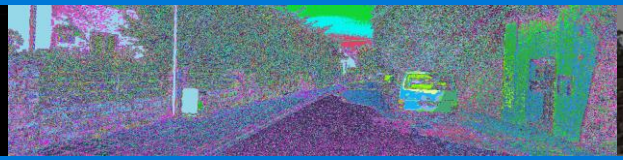


Contextual information

User input

- Vehicle Position and Pose
- Camera Calibration
- Maps

Segmentation



Visual Words

Map prior

Confidence Map

Final Detection

Microsoft

Scene

Semantic Segmentation

Labeling

Priors for Structural Elements of Scene

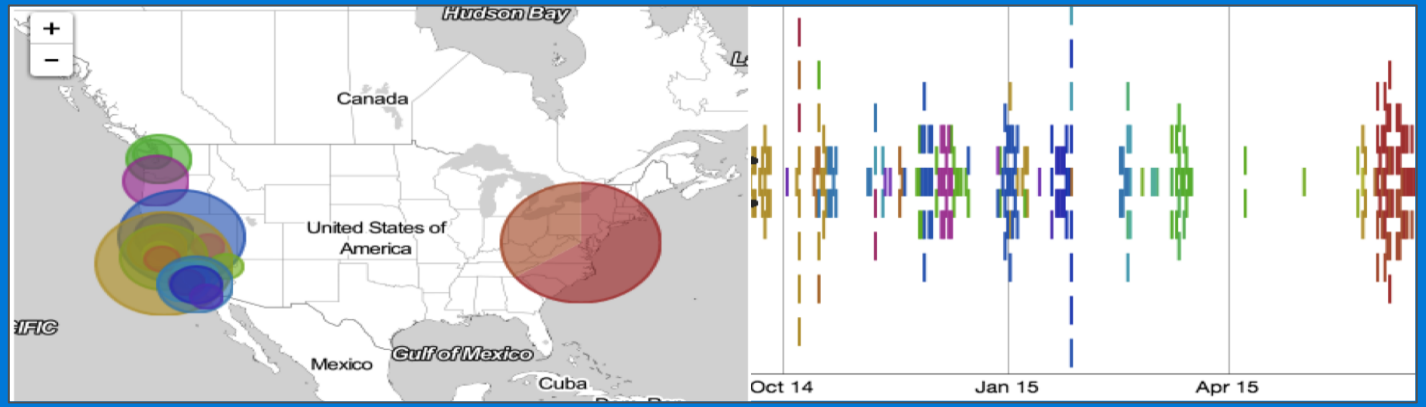
Microsoft

Carnegie Mellon University

MARINUS ANALYTICS

Auton Lab

- Data Mining and Machine Learning applied to find meaningful patterns in publicly available data
- Provides leads and crucial supporting evidence for sex trafficking investigations and victim rescue operations
- Deployed to hundreds of local, state, and federal law enforcement agencies and non-profits across the U.S. and Canada
- Dozens of victims rescued and successful prosecutions made
- Web interface accessible via computers, smart phones, tablets



Example:

Trail of a potential prostitution ring found migrating its operations from the West to the East coast

Find if there is an image of a similar scene

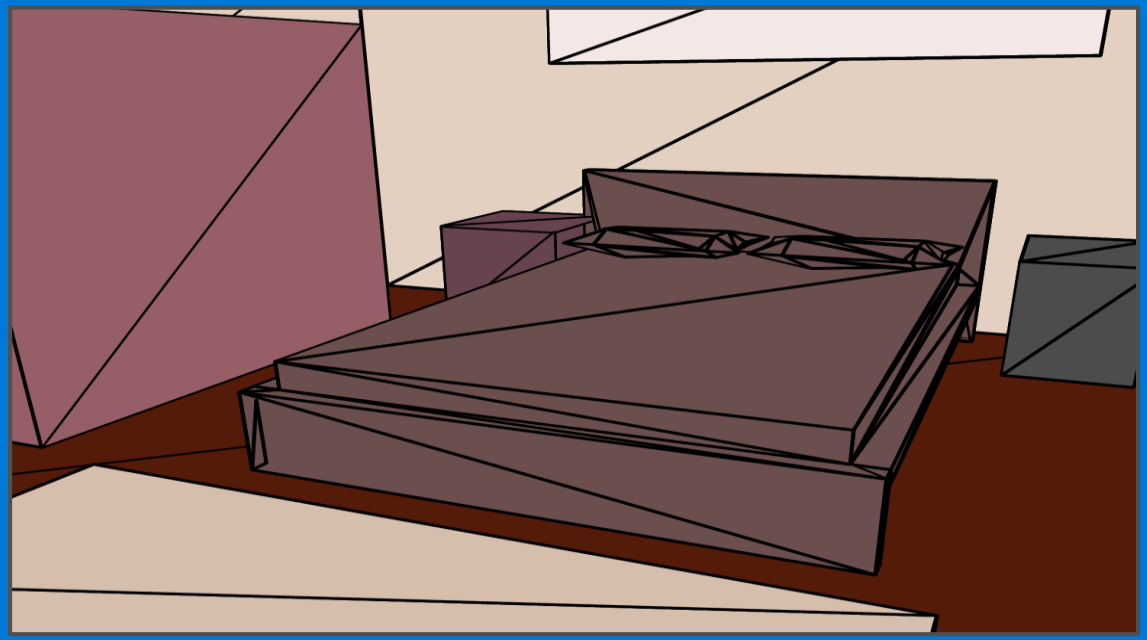
Where was this image taken?

Meritan Apartments Sydney

From which hotel chain?

Sheraton Hotels (North America)

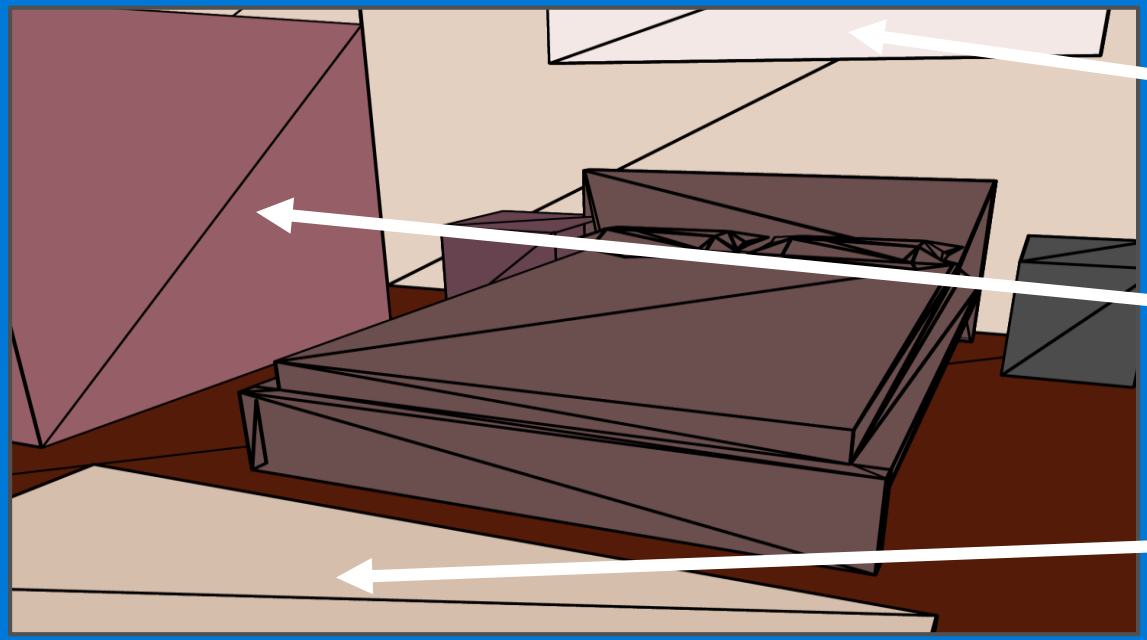
3D Structure

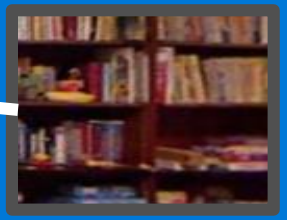


3D Model from Guo and Hoiem, ICCV13.

3D Structure

Style

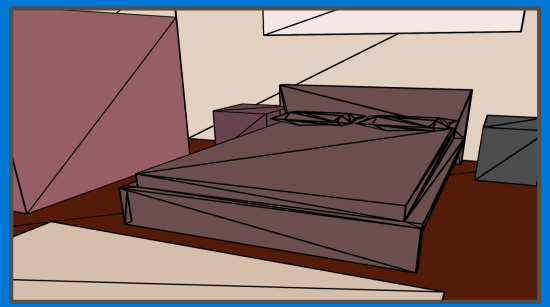




Image

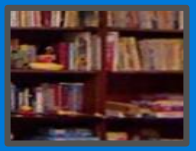
=

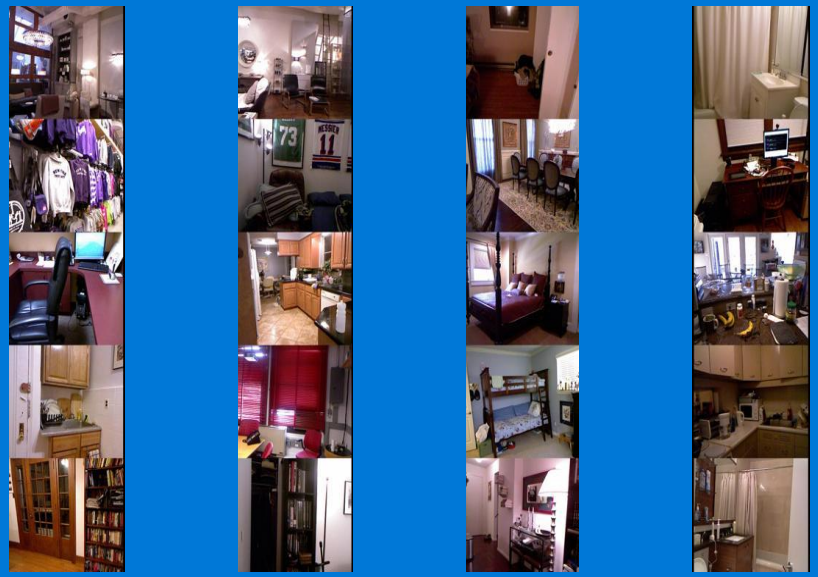
3D Structure



x

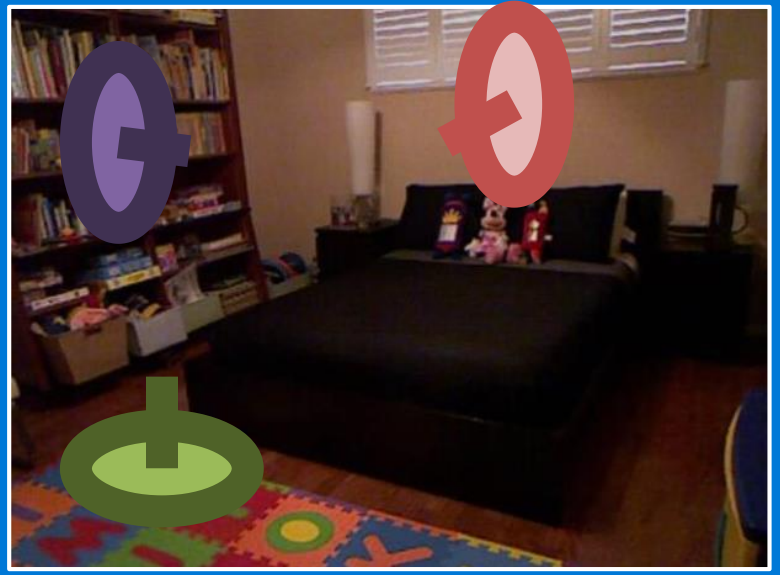
Style





Data

+



Domain knowledge

Element

Detections

Supermarket

Laundromat

Museum

Locker Room

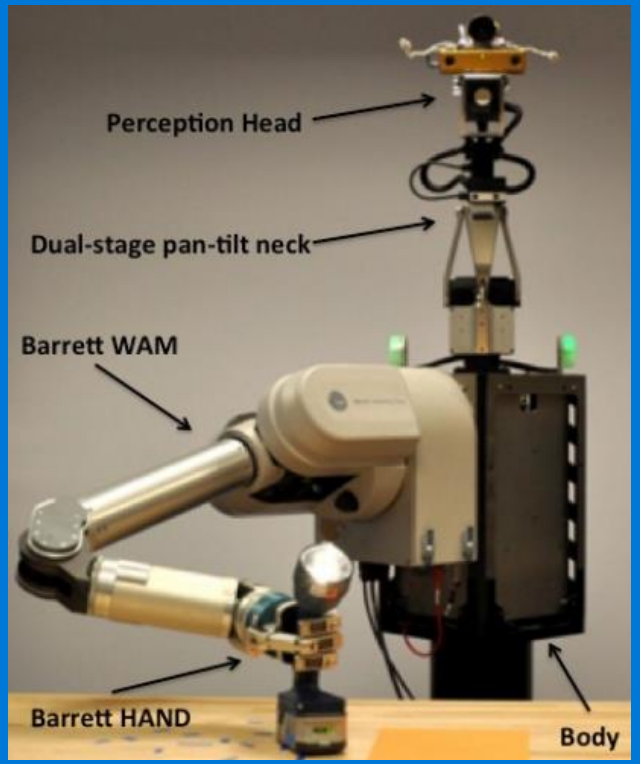
Microsoft

Task-adapted loss function

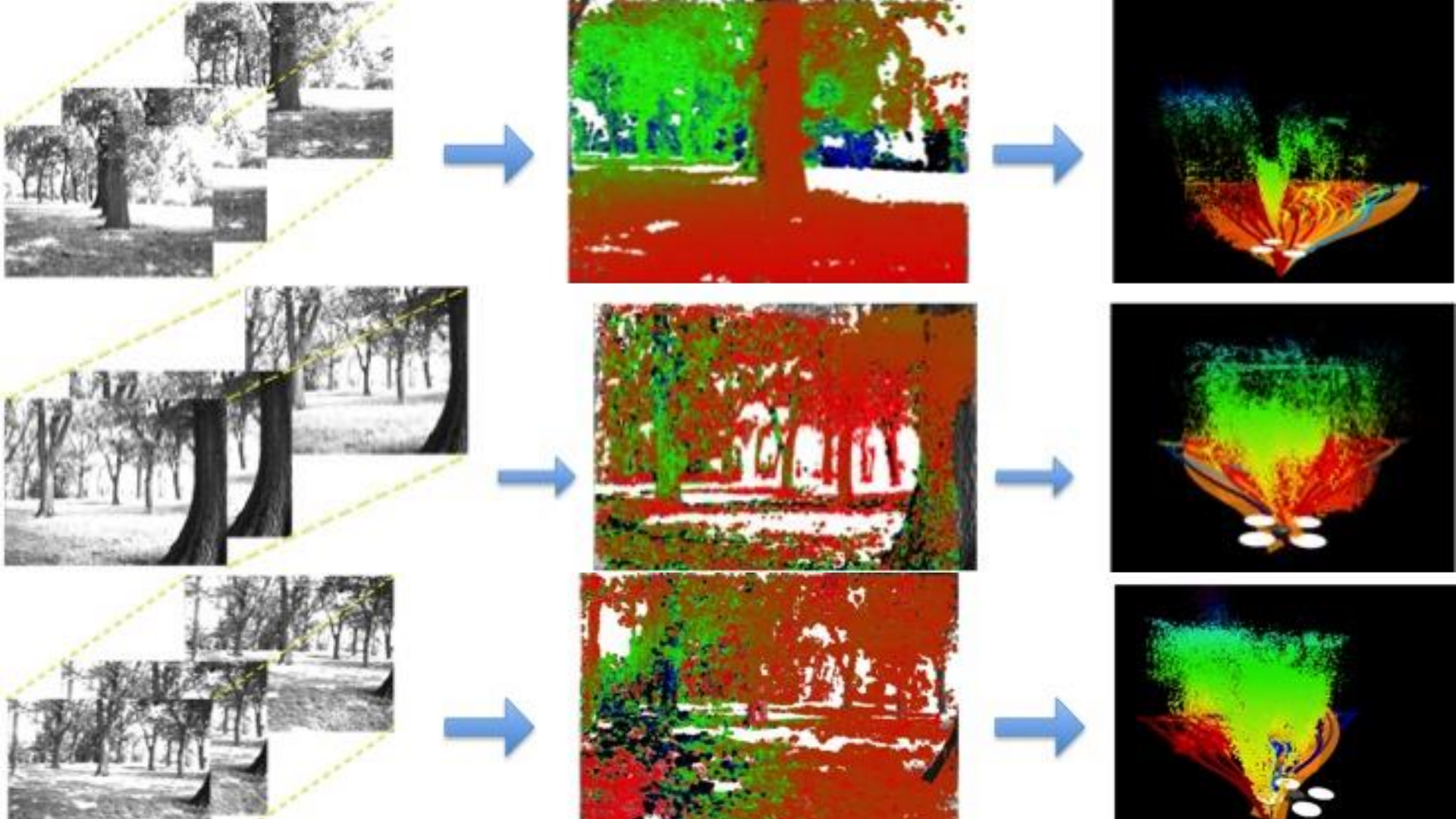

building
building
tree-trunk
person
ground


shrub
building
building
tree-trunk
barrier
person
shrub
ground
small-vehicle

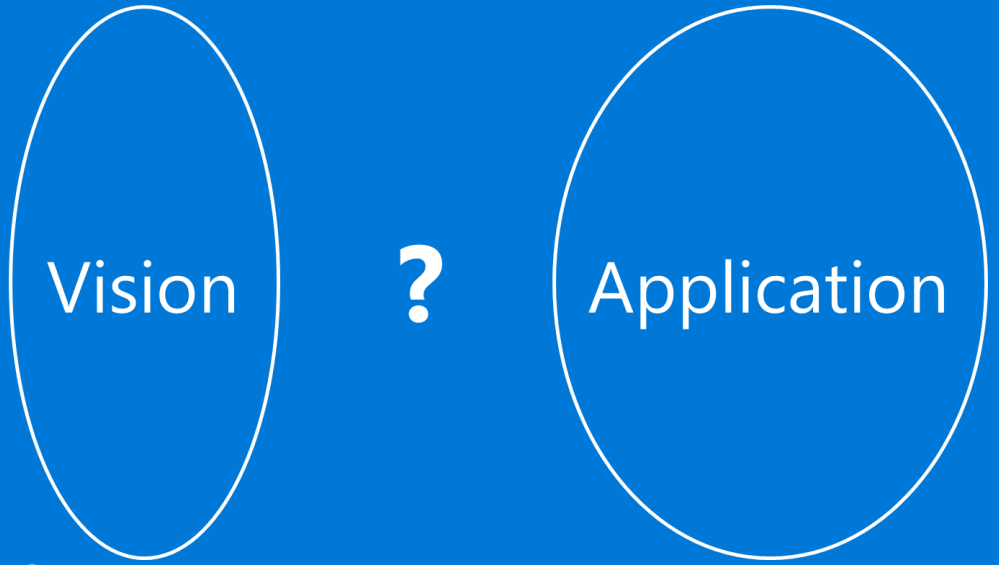
$$Min_{\theta}\ Loss(y(\theta, x), \hat{y})$$


Perception Head
Dual-stage pan-tilt neck
Barrett WAM
Barrett HAND
Body


Stay to the right of the car; screen the back of the building that is behind the car.



- Multiple hypotheses
- Input filtering
- Knowledge
- Task-based loss
- .....
- Anytime prediction
- Variable depth resolution
- Small sample adaptation
- Distributed computation
- .....



- Jean Oh, Arne Suppe, Luis Navarro-Serment, Drew Bagnell, Debadeepta Dey, Shreyansh Daftry, Devi Parikh, Ali Farhadi, Peng Zhang, Jiuling Wang, Ankit Laddha, Abhinav Gupta, David Fouhey