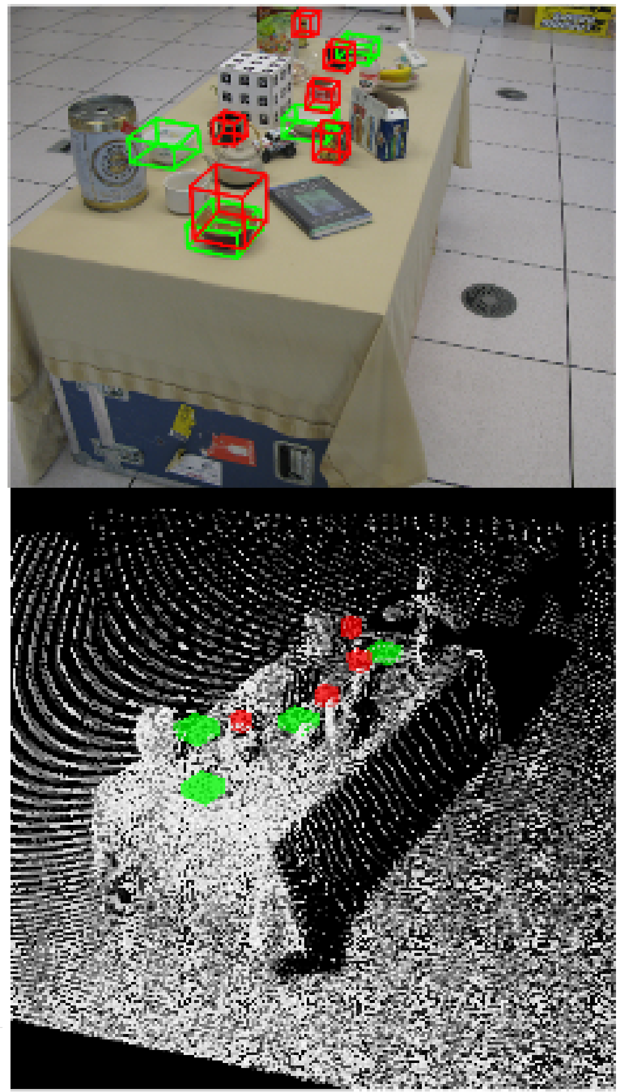
2 Problems in Robot Vision: Recognizing objects and visual localization

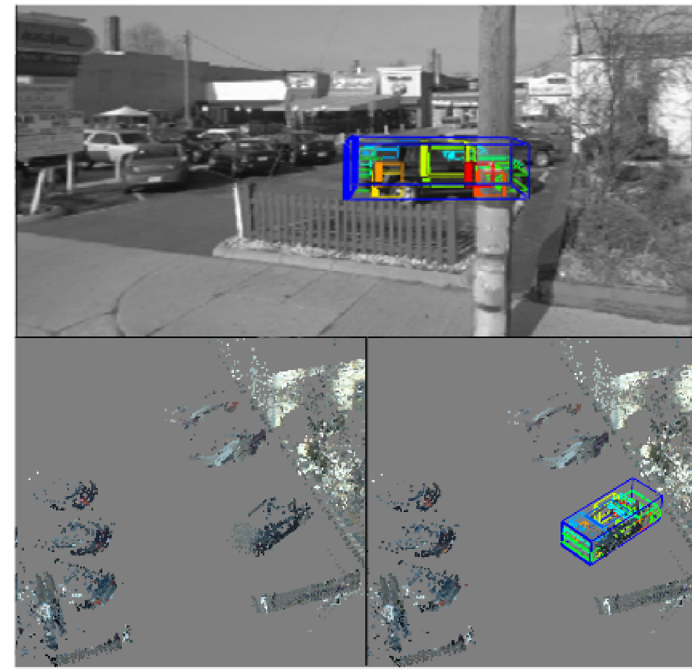
Guest Lectures by David Meger
McGill CS 417
November 4 & 6, 2013

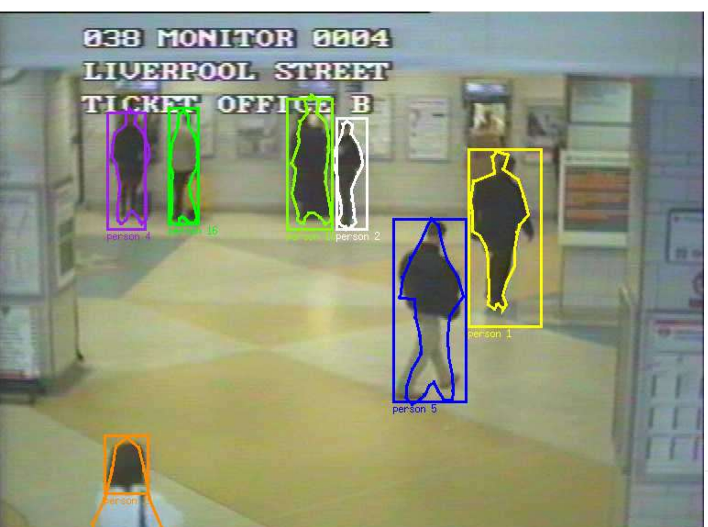
McGill

Motivation

Problem 1: Object recognition







McGill

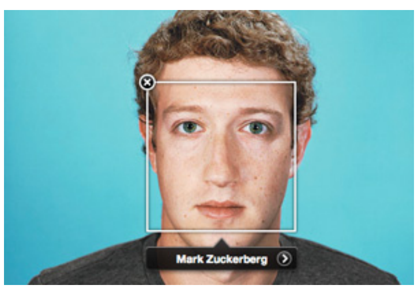
What is object recognition?

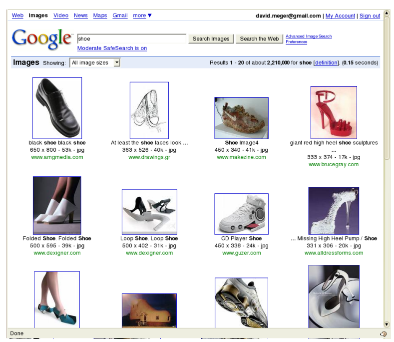
- Input:
  - Training information about objects provided by users or the system designer
  - Perceptual data: Ideally just a single image, maybe a point cloud, video, user assistance, etc
- Output:
  - Object labels: Ideally accurate match to what you were told in training. Maybe includes "ontological" information.
  - Object locations: Ideally 3D, but maybe 2D. Maybe includes pose.

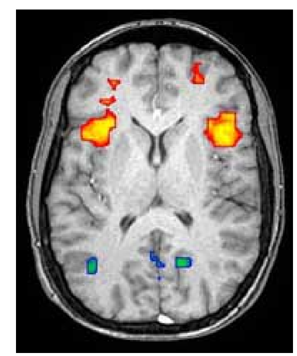
Organizations that want to recognize objects

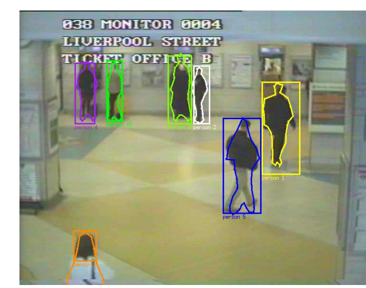
Why?

Organizations that want to recognize objects





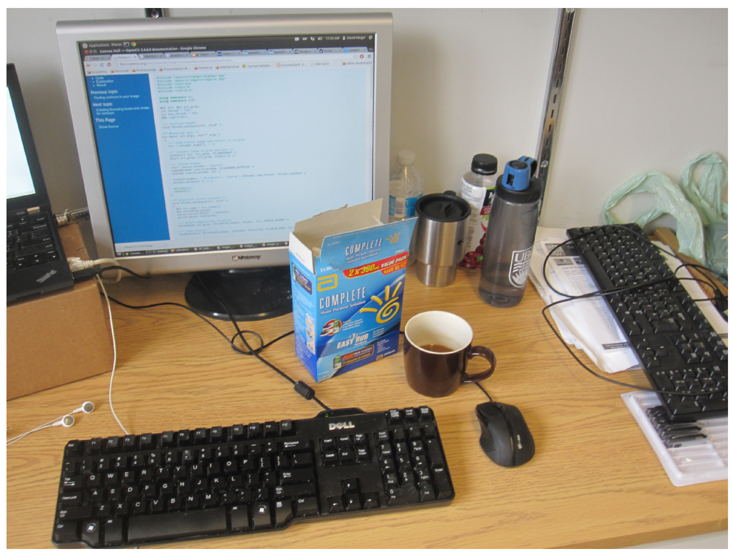




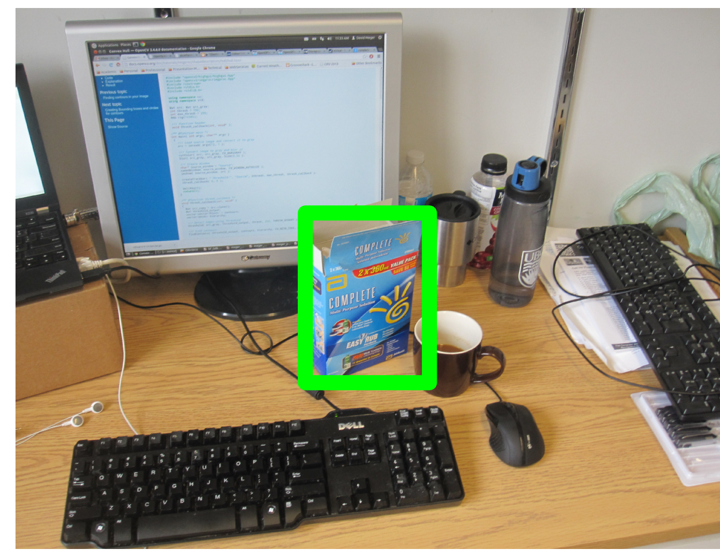
How?

User: This is contact lens solution

User: What objects are here?



System: Here is the contact lens solution!



McGill

Example Solutions

1) Template Matching: Find the area of an image that matches examples as exactly as possible
2) Shape Matching: Perform the same operation only on "edges"
3) Color matching: Forget image positions, only consider R,G,B description

Challenges of Image Recognition on Realistic Images

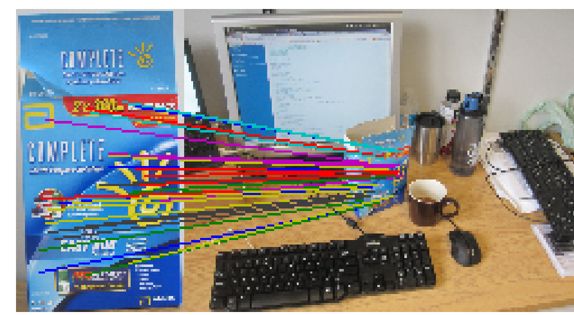
- Lighting
- Scale
- Position
- Orientation
- Projection
- Deformation and articulation
- Occlusion

Solution: local image features

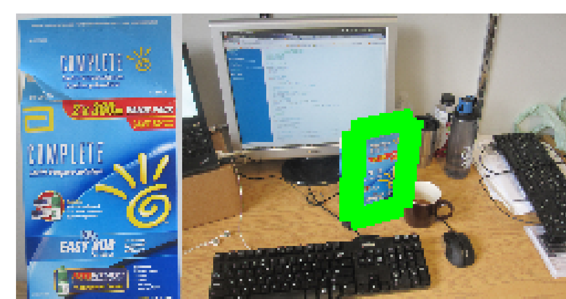
- Some local patches are "recognizable"
- Feature detectors find repeated locations
- Feature descriptors provide invariant local representations
- Groups of features that agree on geometry are kept as object guesses

137 tentative matches

35 final matches



Estimated Object

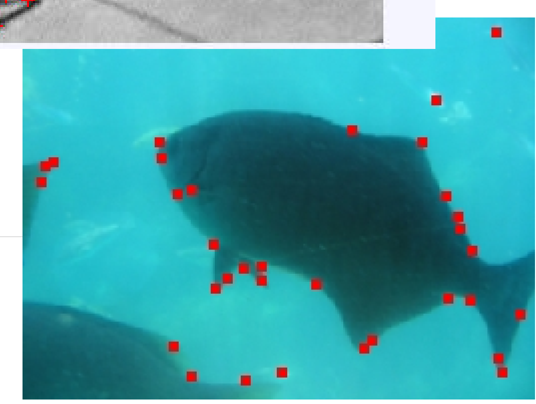


What makes a good local feature?

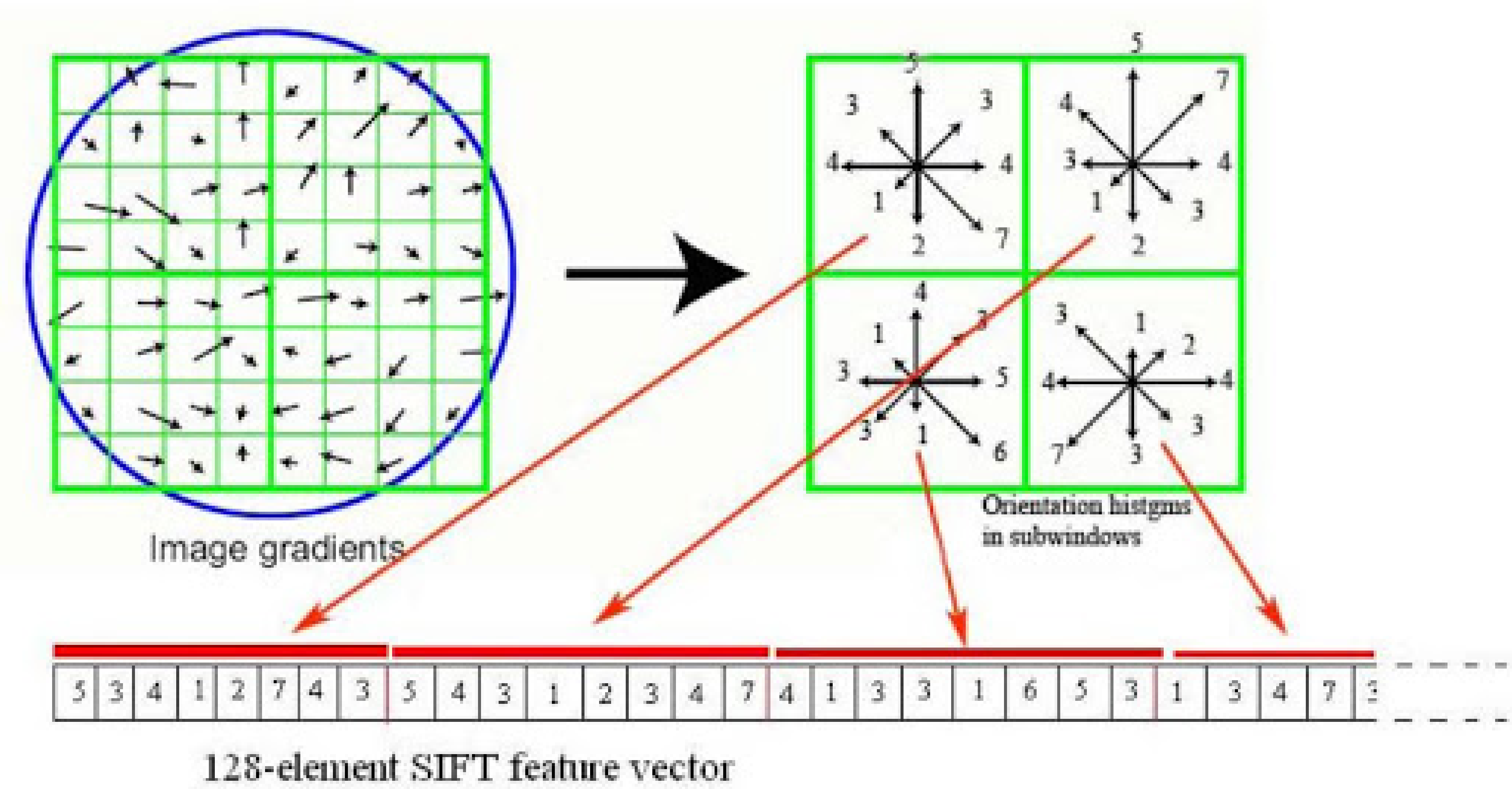
- High Recall: They can be located again
- High Precision: They are not too confused with other patches
- Several Alternative Examples:
  - Harris Corners
  - SURF
  - SIFT
  - FREAK
  - BRIEF
  - etc

Harris Corners

- Operation based on strong gradients in multiple directions
  - Not edges, those can not be reliably located
  - Not flat regions, those cannot be described easily



Scale Invariant Feature Transform Descriptors



McGill

Feature Matching Algorithm

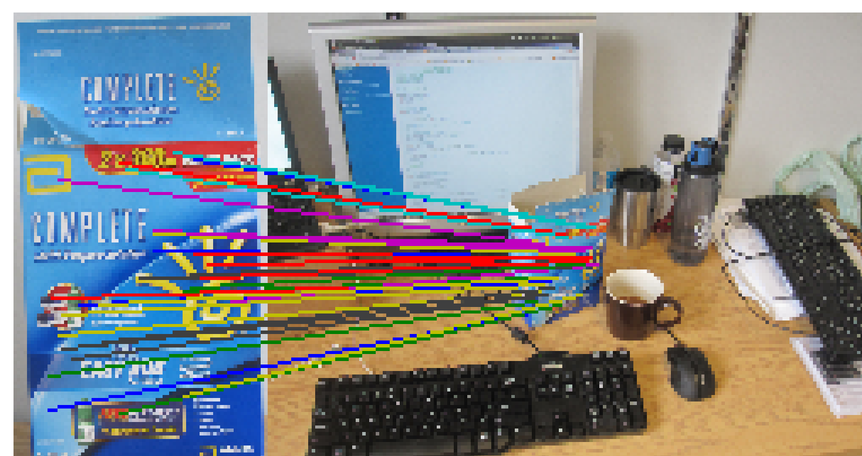
- For each feature, find the nearest neighbor (closest 128 dimensional descriptor)
- If matches are not "bi-directional", discard
- Threshold on ratio of best match score to 2nd best match score
- Everything remaining is a candidate match to check geometry

Geometric Consistency

- Features on object will be a sub-set of all matches, with much noise
- Idea: search for a large set of features that agree on geometry
- Algorithm: Randomized Sampling and Consensus (RANSAC)

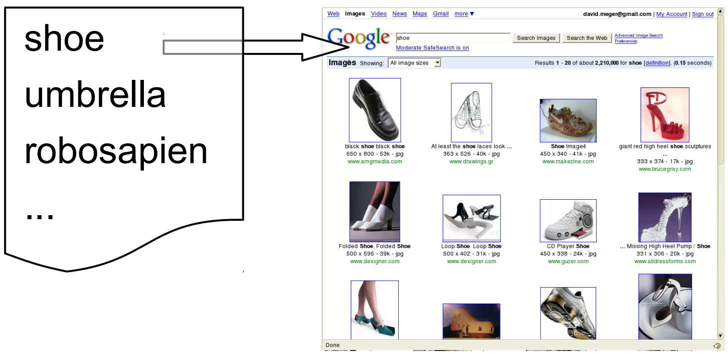
137 tentative matches

35 final matches



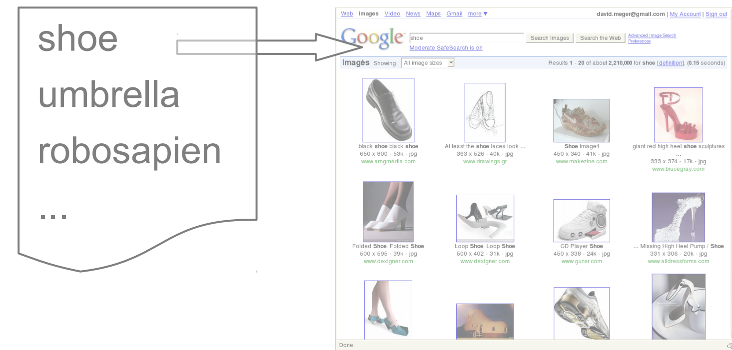
Example of Live Robot Recognition: SRVC Contest

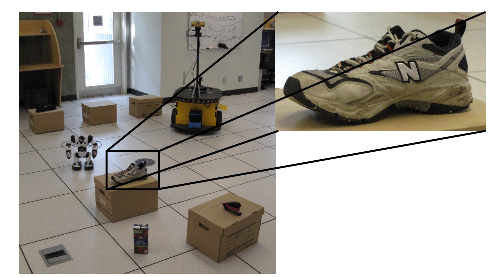
- Phase 1: Automated analysis of web to learn previously unseen objects



SRVC Contest

- Phase 1: Automated analysis of web to learn previously unseen objects
- Phase 2: Explore contest environment, collect imagery

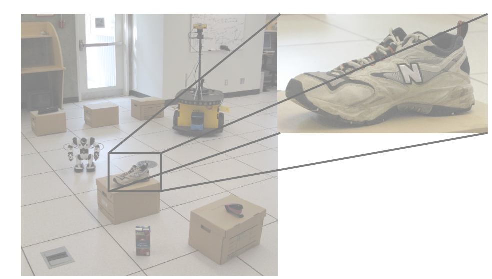




SRVC Contest

- Phase 1: Automated analysis of web to learn previously unseen objects

- Phase 2: Explore contest environment, collect imagery

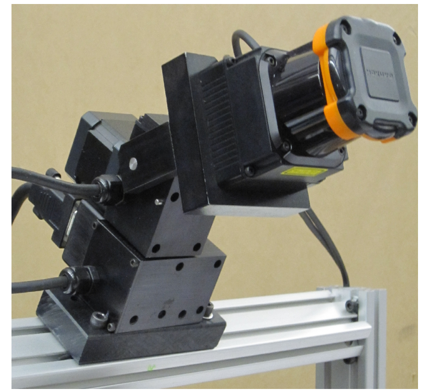


- Phase 3: Perform recognition

= shoe

Curious George Robot

<u>Powerbot mobile base with sensor tower:</u> provides excellent indoor autonomy, long battery life, numerous sensors and ability to carry very significant computation on-board



<u>Tilting laser rangefinder:</u> captures >180 degree 3D scans of the environment

<u>High-res digital camera:</u> adapts zoom and grabs 10MP images with consumer-grade quality

<u>Bumblebee stereo camera:</u> high frame-rate visual information and narrow field-of-view depth sensing

McGill

Contest Environment

McGill

(a)

(b)

(c)

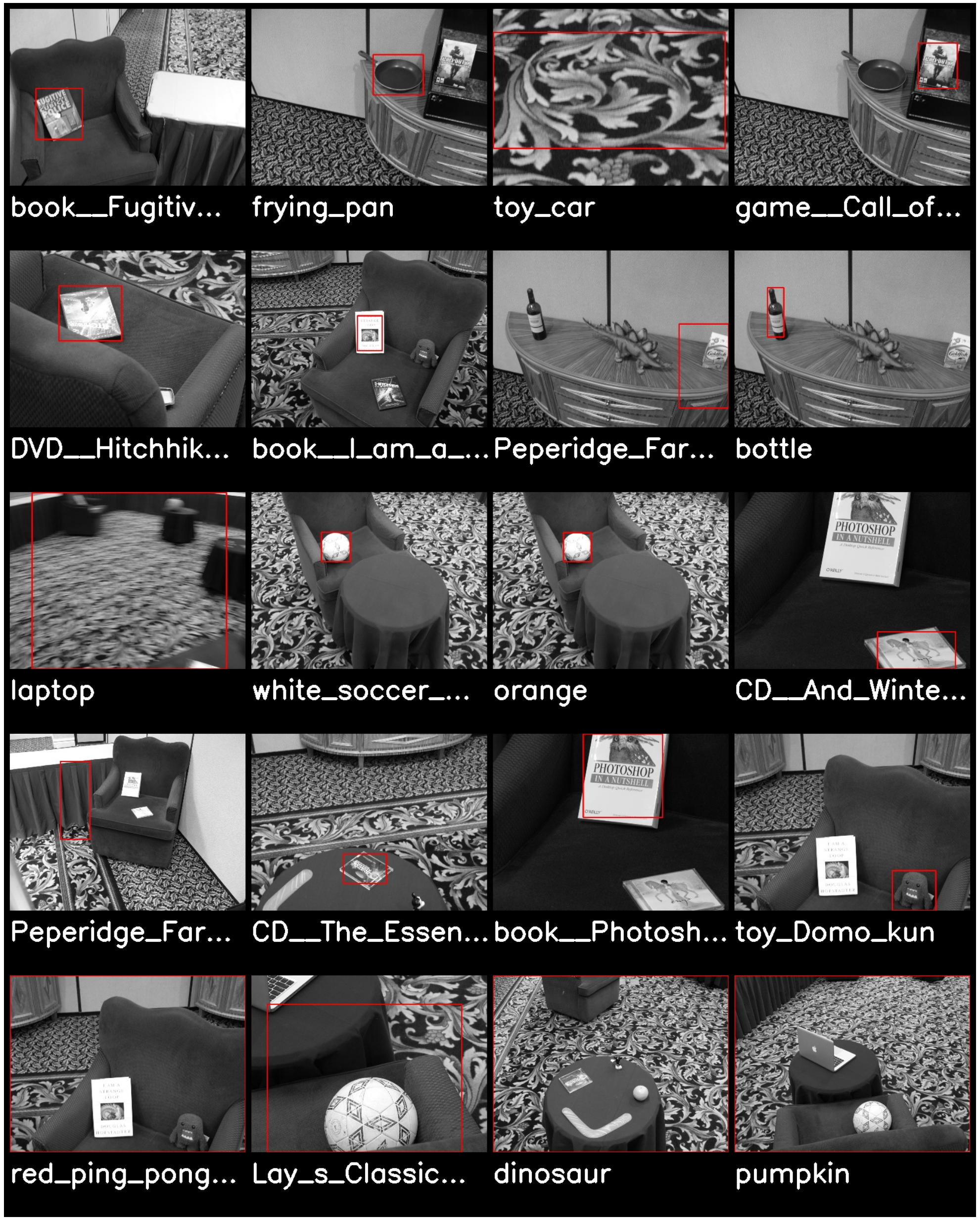
(d)

(e)

(f)

Curious George Results

- Scoring:
  - 8 of 12 specific instances
  - 4 of 8 generic categories



Result Summary

Year	Type	Name	Result
2009	category	pumpkin	incorrect
		orange	correct
		red ping pong paddle	incorrect
		white soccer ball	incorrect
		laptop	incorrect
		dinosaur	incorrect
		bottle	correct
		toy car	incorrect
		frying pan	correct
	instance	book “I am a Strange Loop” by Douglas Hofstadter	correct
		book “Fugitive from the Cubicle Police”	incorrect
		book “Photoshop in a Nutshell”	correct
		CD “And Winter Came” by Enya	correct
		CD “The Essential Collection” by Karl Jenkins *et al.*	correct
		DVD “Hitchhiker’s Guide to the Galaxy” widescreen	correct
		game “Call of Duty	correct
		toy Domo	correct
		Lay’s Classic Potato Chips	correct
		Peperidge Farms Goldfish Baked Snack Crackers	correct
		Peperidge Farm Milano Distinctive Cookies	correct

Major challenges?

1)Safety

2)Exploration

3)Control of the camera to get useful images

4)Comparing web images to the robot's

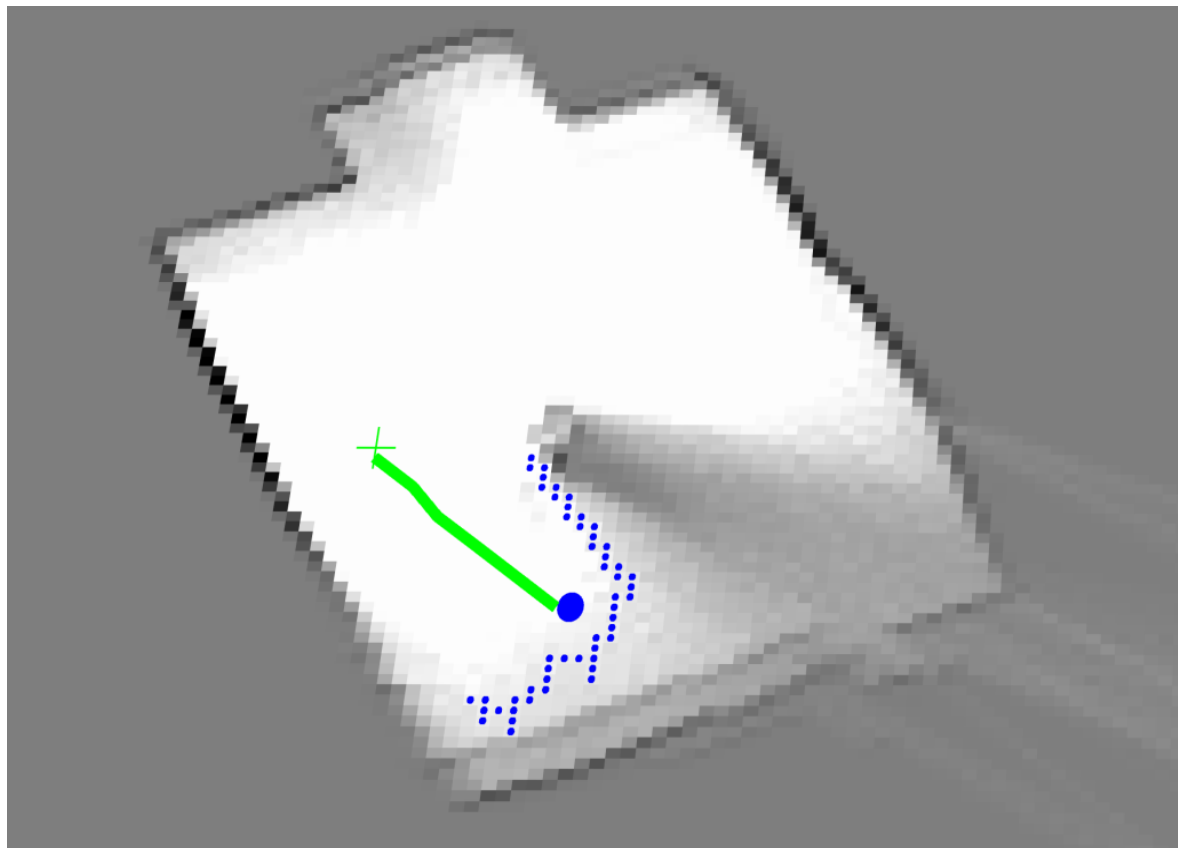
5)Modeling a pumpkin

Navigation Approaches

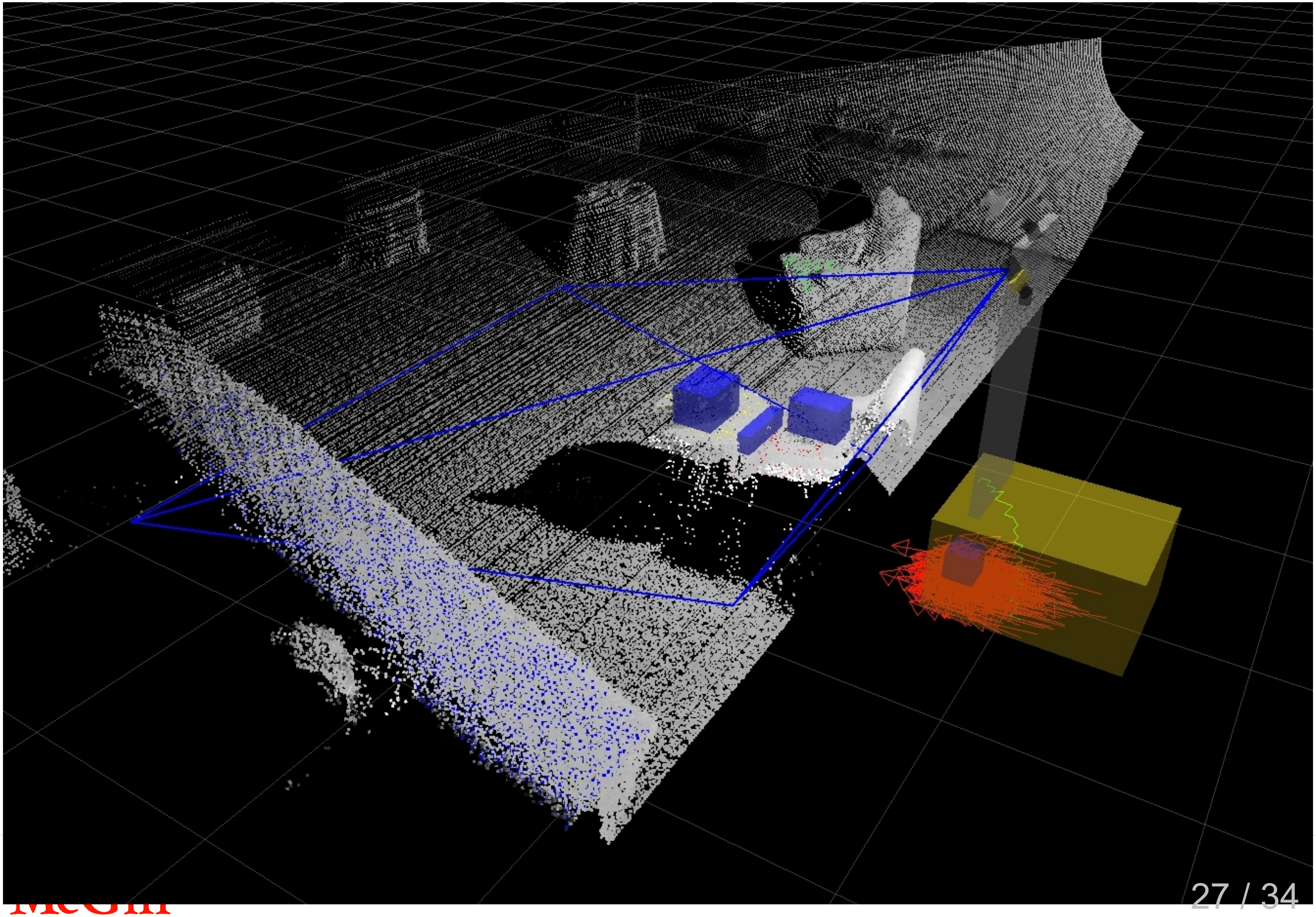
- 2D occupancy grids for safe and effective navigation
- Video
- How to relate these maps to visual tasks?
- What type of algorithms are needed to provide the answers?

Area Coverage

- An uncertain map boundary is detected, and a plan to explore the **frontier** is executed (Yamauchi CIRA97)

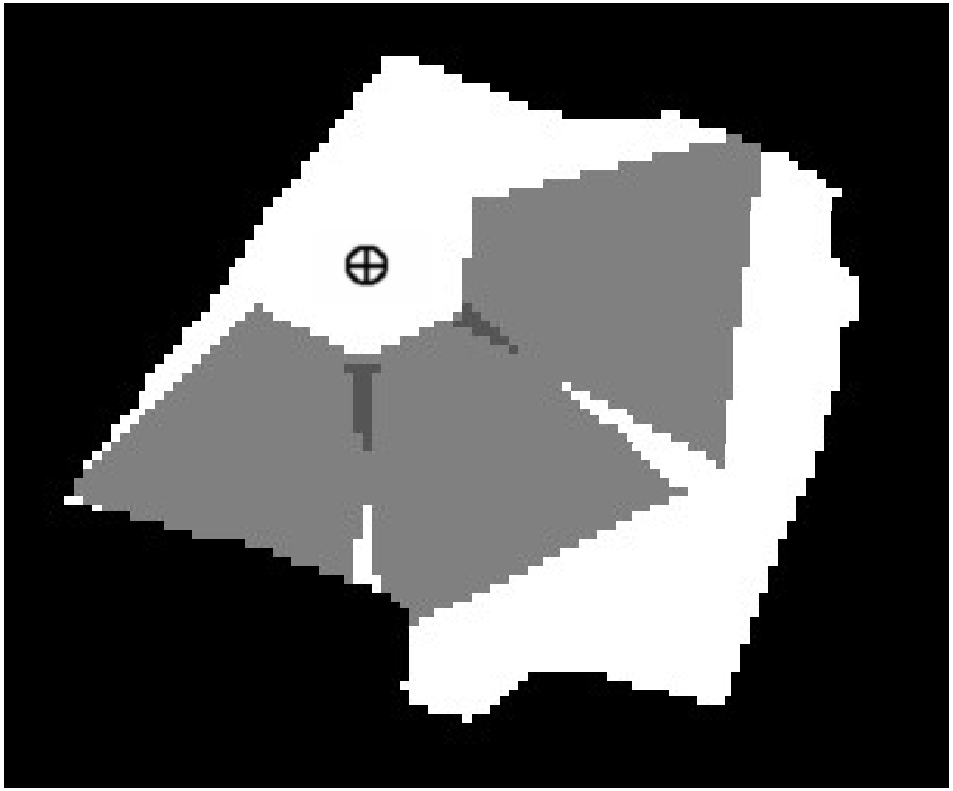


Visual Coverage



Visual Coverage

- Use known robot model to determine area captured with each image. Plan to achieve full coverage.

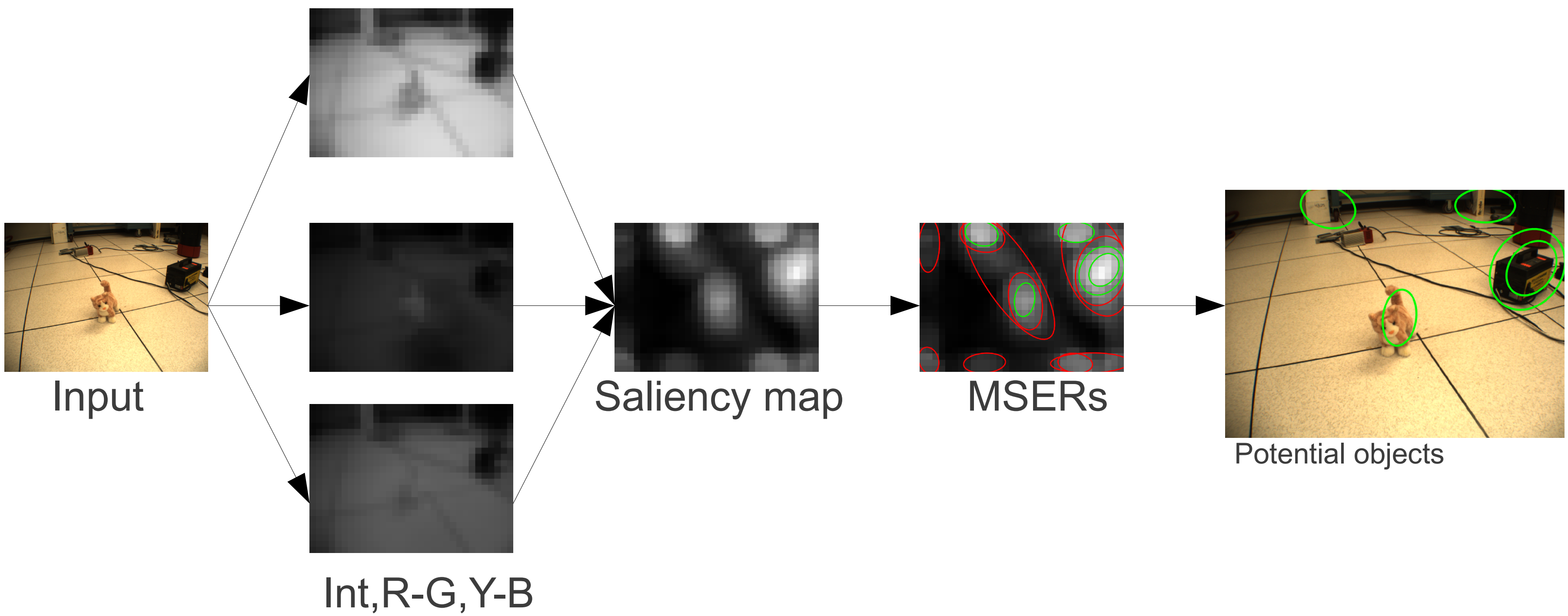


Which pictures to take?

- Humans “frame” pictures to capture the desired content at the desired angle, focus, etc
- How can we replicate that with a robot?

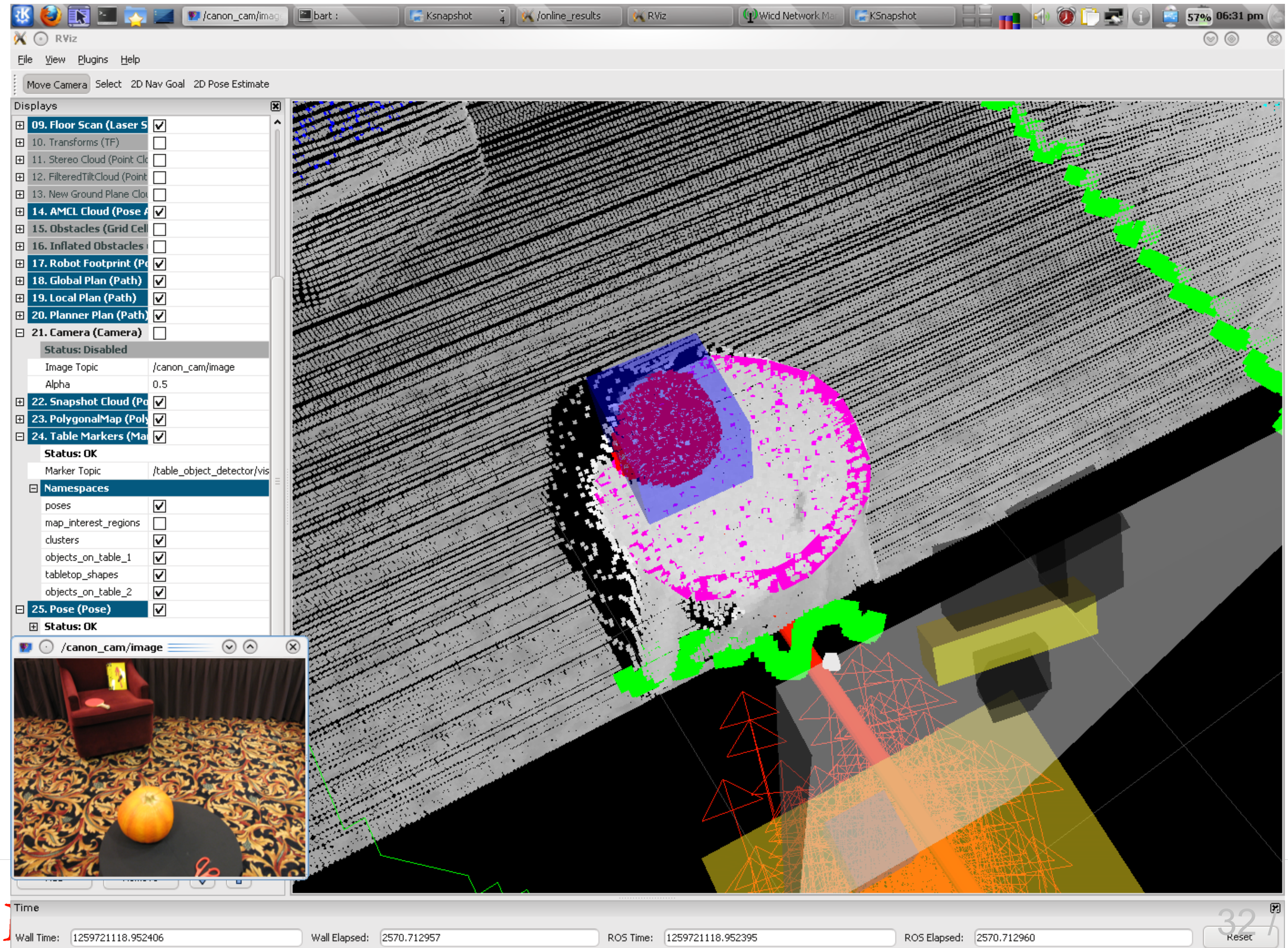
Visual Saliency

- One way for intelligent systems to focus on useful content based on analysis of “interesting-ness”

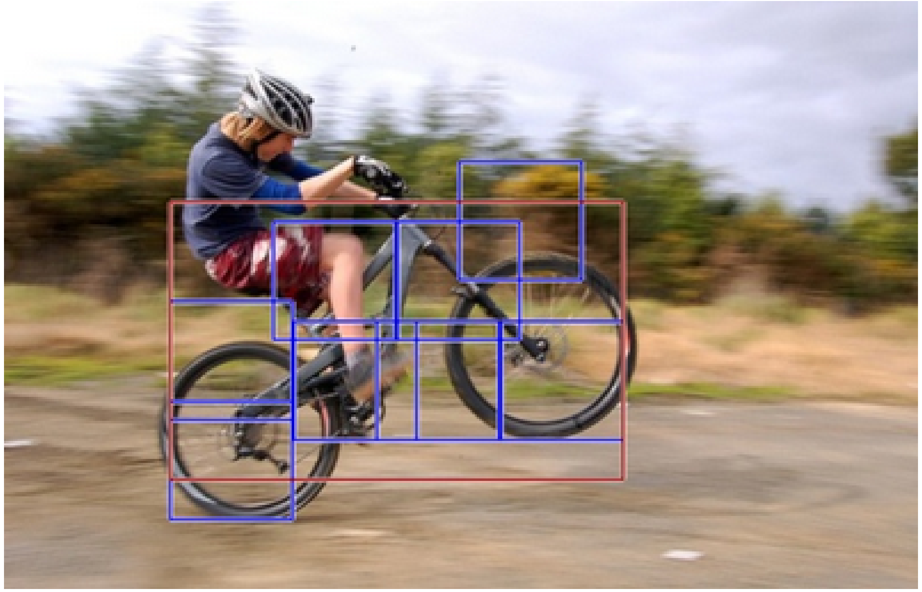


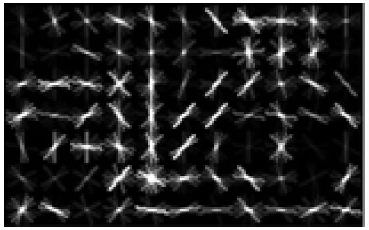
Multi-Scale Saliency

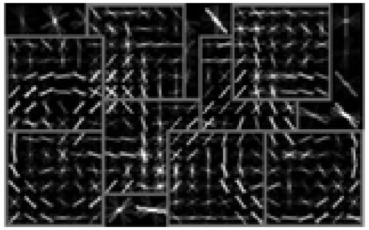
Using 3D Data for attention

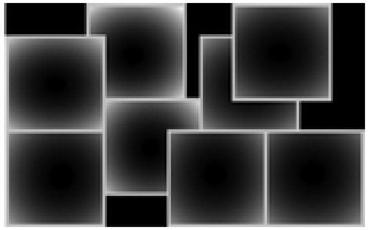


Parts-Based Detection









More info at:
http://people.cs.uchicago.edu/~rbg/latent/

McGill

Questions