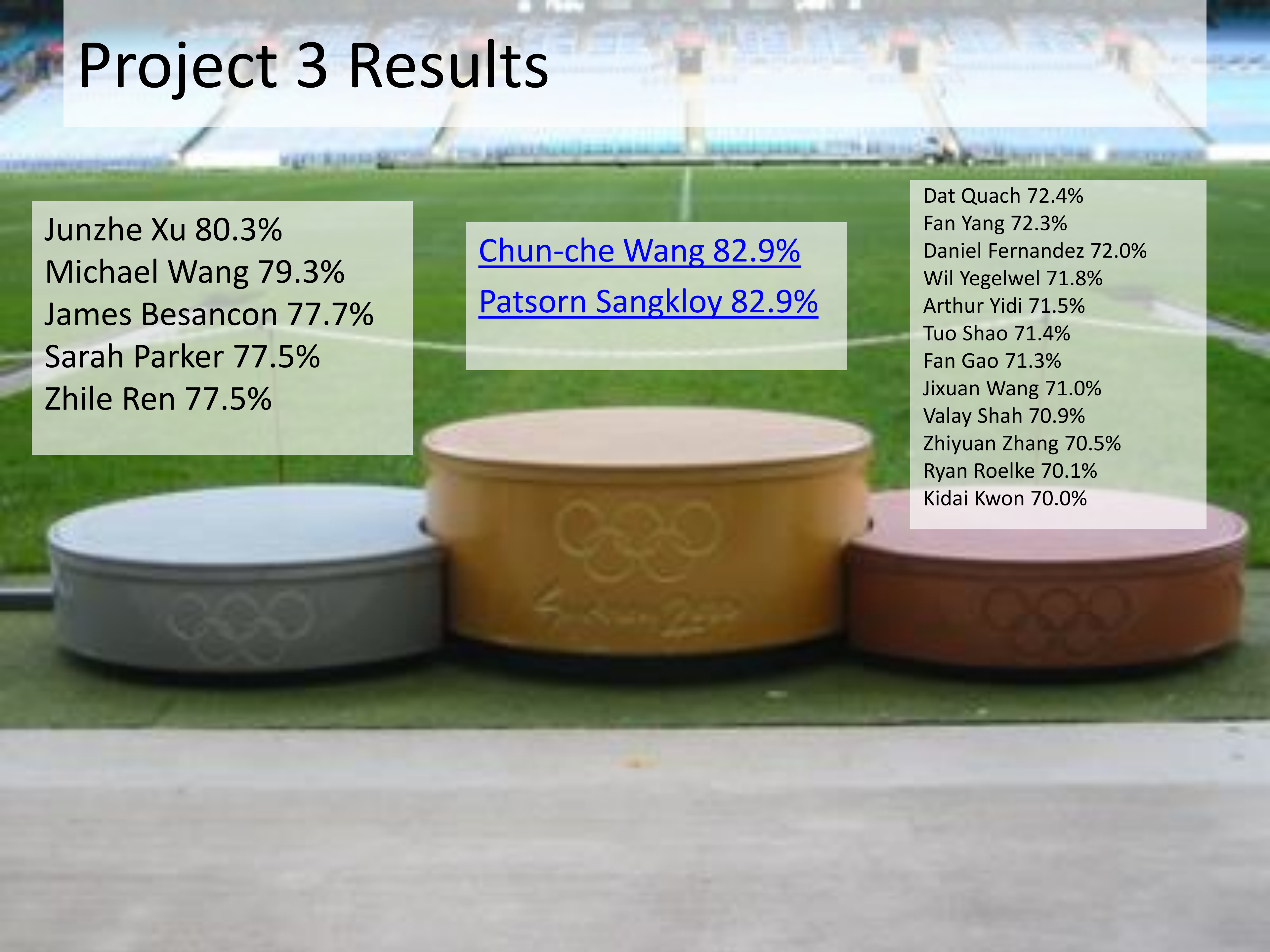
Project 3 Results



Junzhe Xu 80.3%
Michael Wang 79.3%
James Besancon 77.7%
Sarah Parker 77.5%
Zhile Ren 77.5%

Chun-che Wang 82.9%
Patsorn Sangkloy 82.9%

Dat Quach 72.4%
Fan Yang 72.3%
Daniel Fernandez 72.0%
Wil Yegelwel 71.8%
Arthur Yidi 71.5%
Tuo Shao 71.4%
Fan Gao 71.3%
Jixuan Wang 71.0%
Valay Shah 70.9%
Zhiyuan Zhang 70.5%
Ryan Roelke 70.1%
Kidai Kwon 70.0%

Context and Spatial Layout

Computer Vision

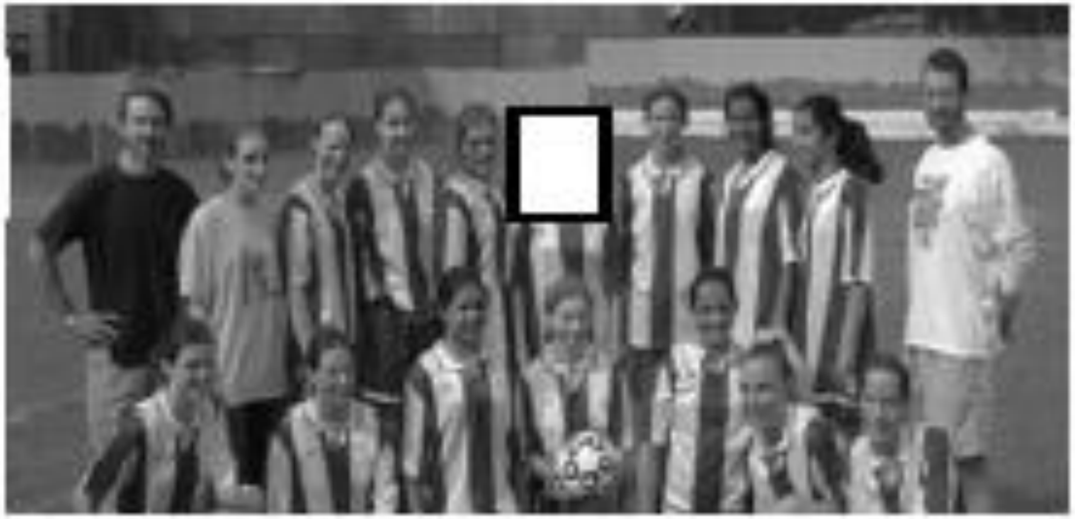
CS 143, Brown

James Hays

Many Slides from Derek Hoiem and Antonio Torralba

Context in Recognition

- Objects usually are surrounded by a scene that can provide context in the form of nearby objects, surfaces, scene category, geometry, etc.



Contextual Reasoning

- Definition: Making a decision based on more than *local* image evidence.

Context provides clues for function

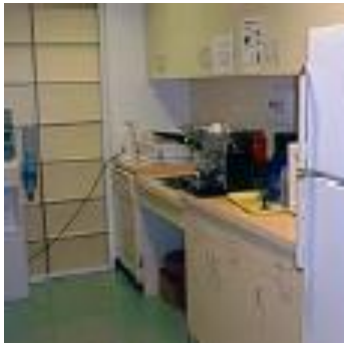
- What is this?

These examples from Antonio Torralba

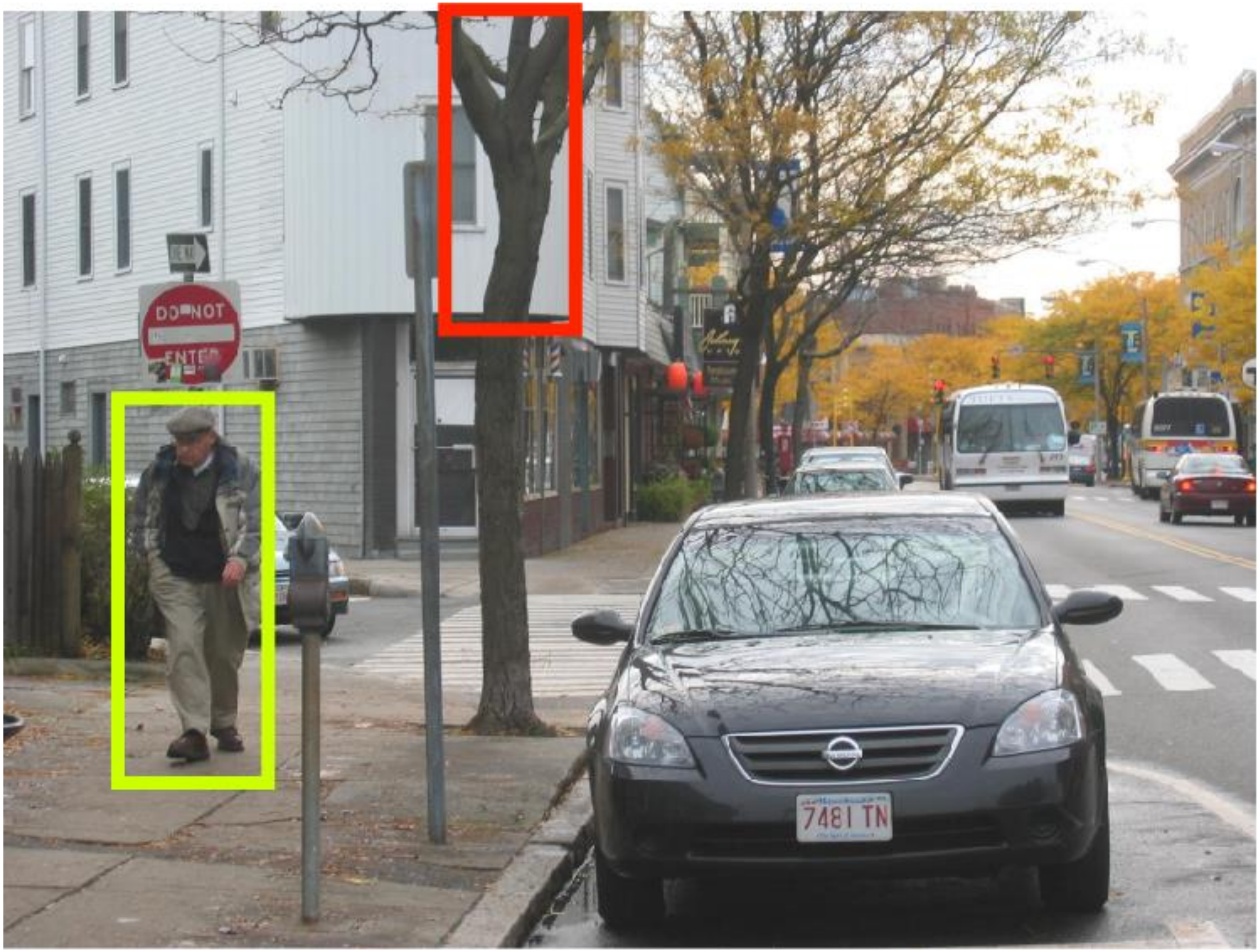
Context provides clues for function

- What is this?

- Now can you tell?

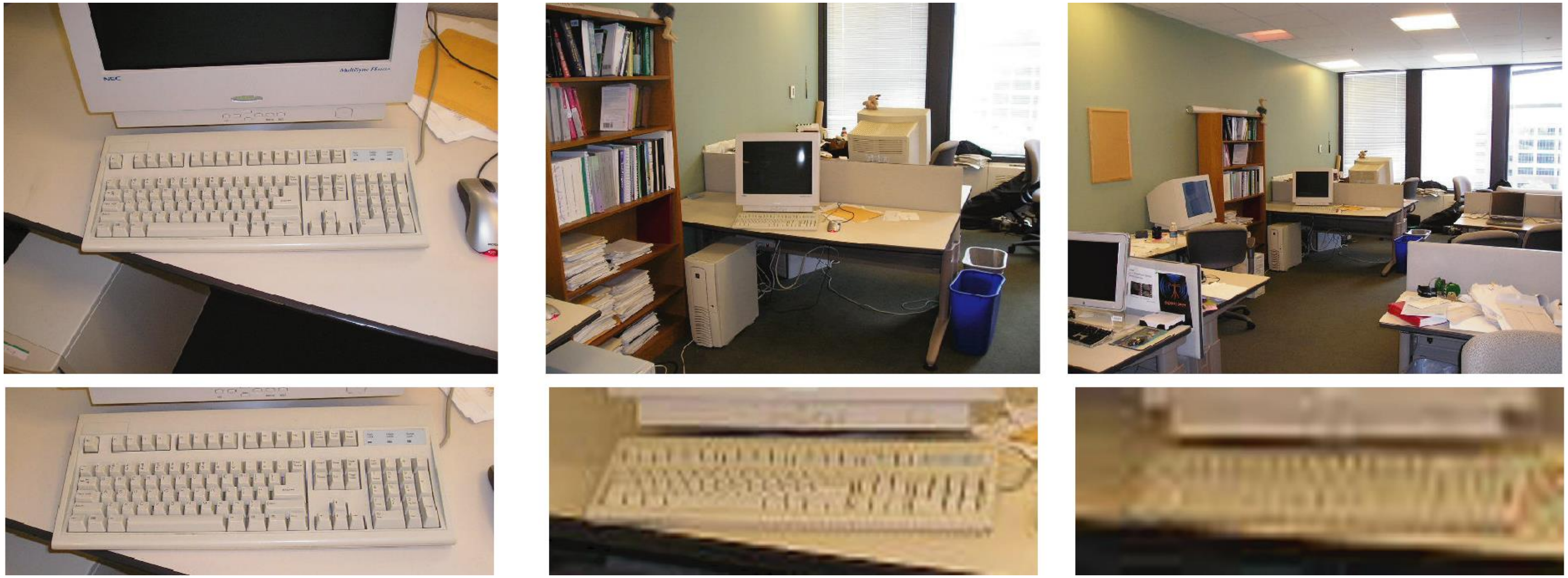


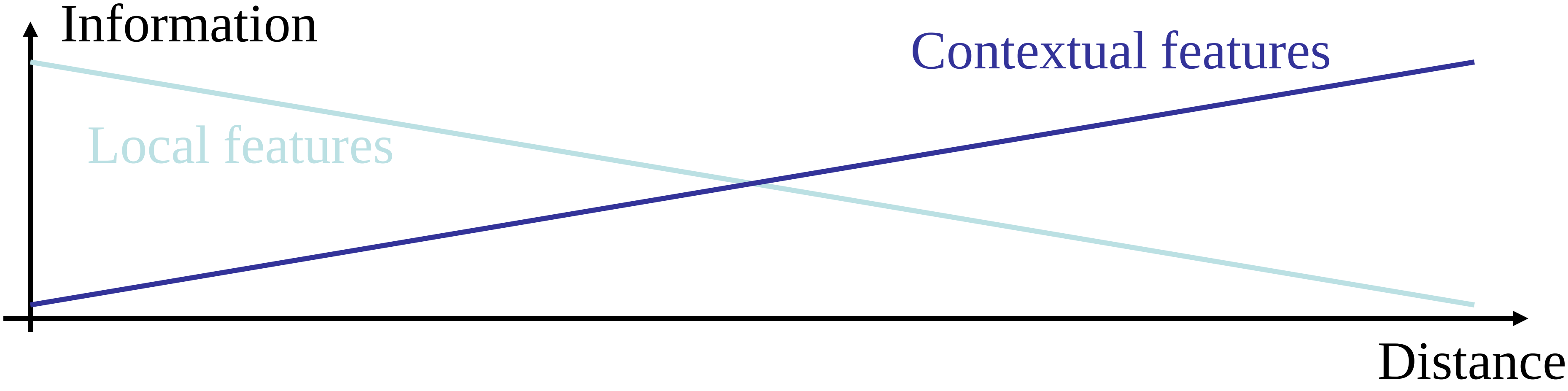
Is local information enough?



Is local information even enough?

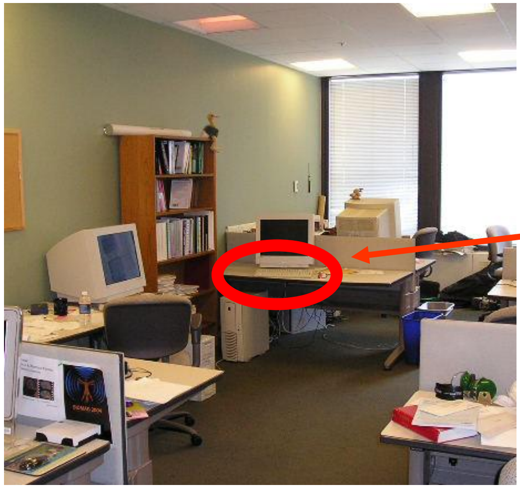
Is local information even enough?



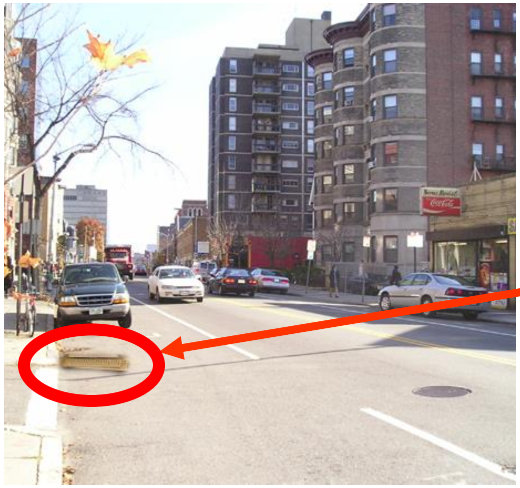


The system does not care about the scene, but we do…

We know there is a keyboard present in this scene even if we cannot see it clearly.

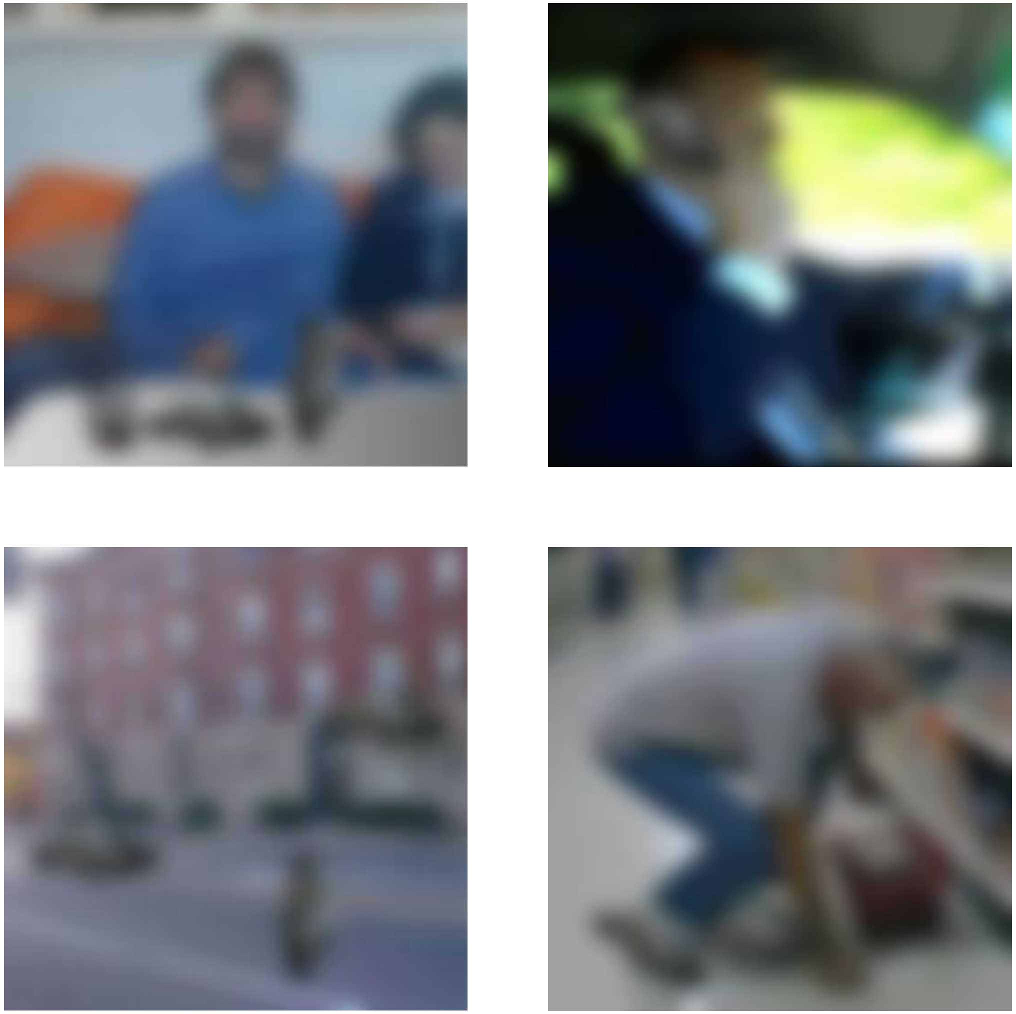


We know there is no keyboard present in this scene

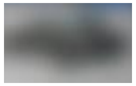


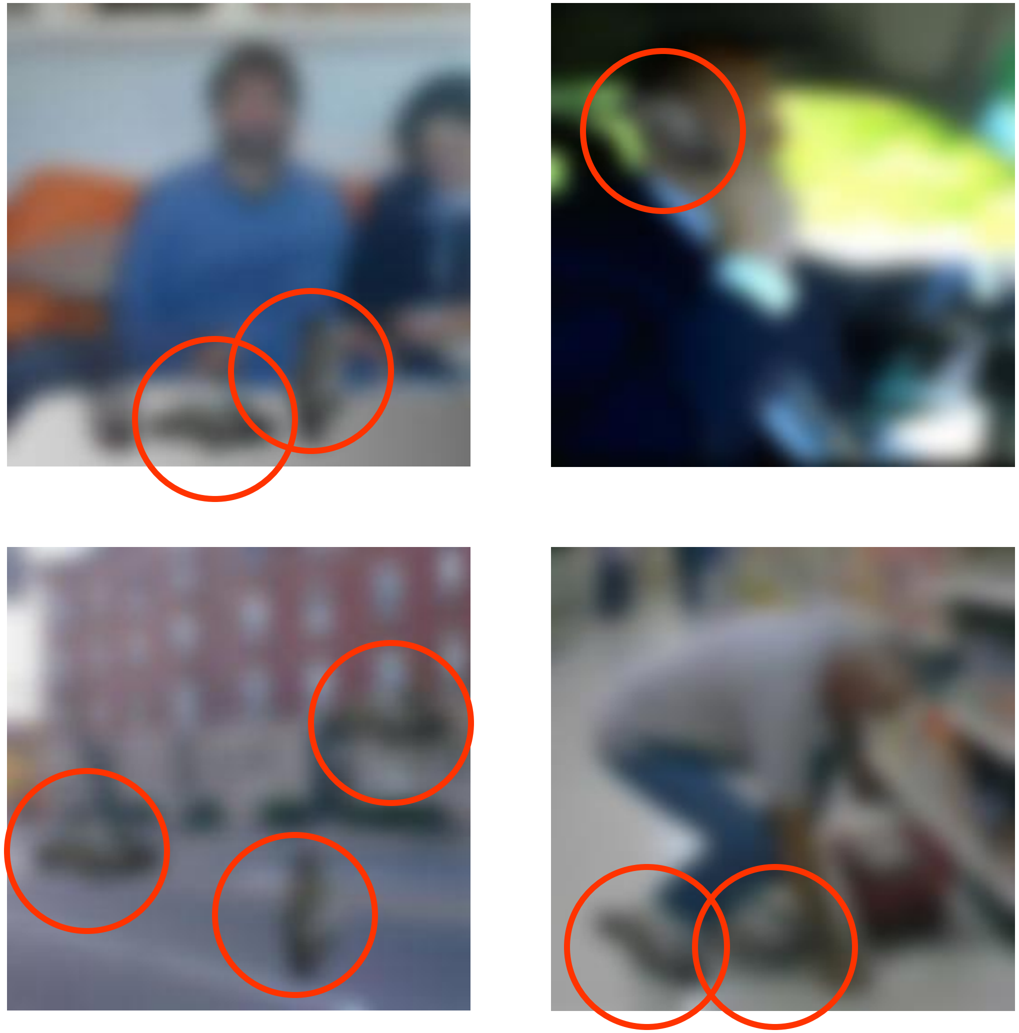
… even if there is one indeed.

The multiple personalities of a blob



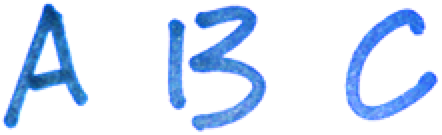
The multiple personalities of a blob



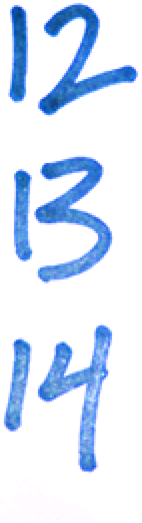


A B C

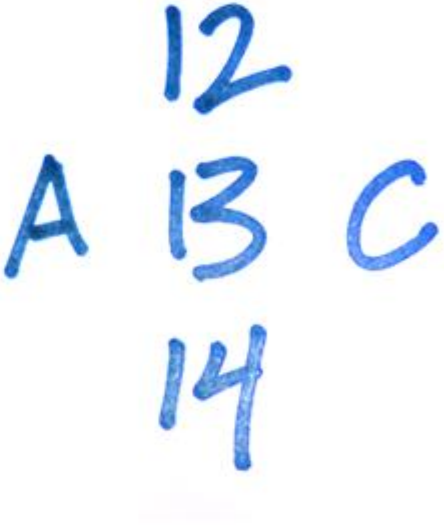
12
13
14



A B C

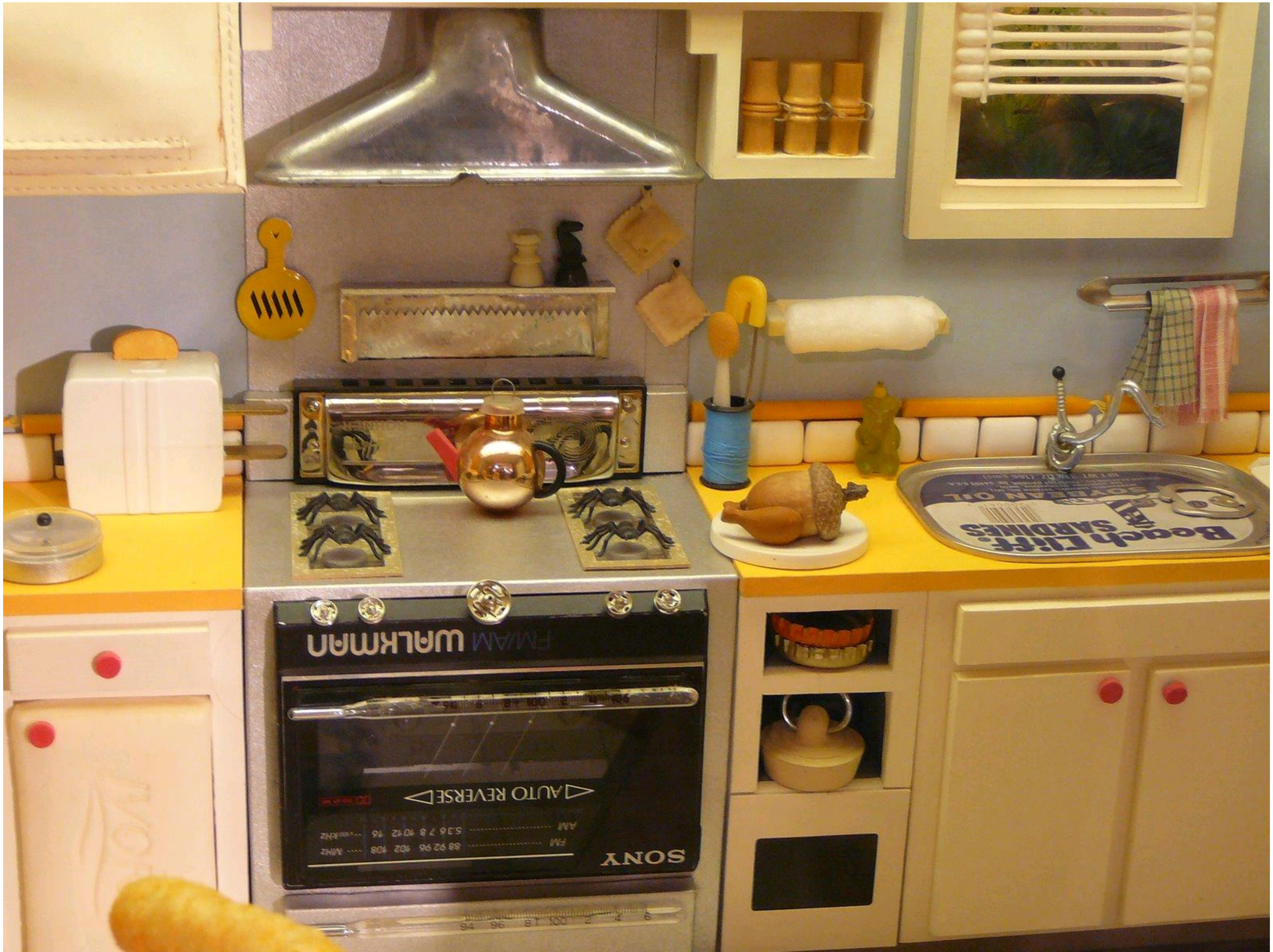


12
13
14

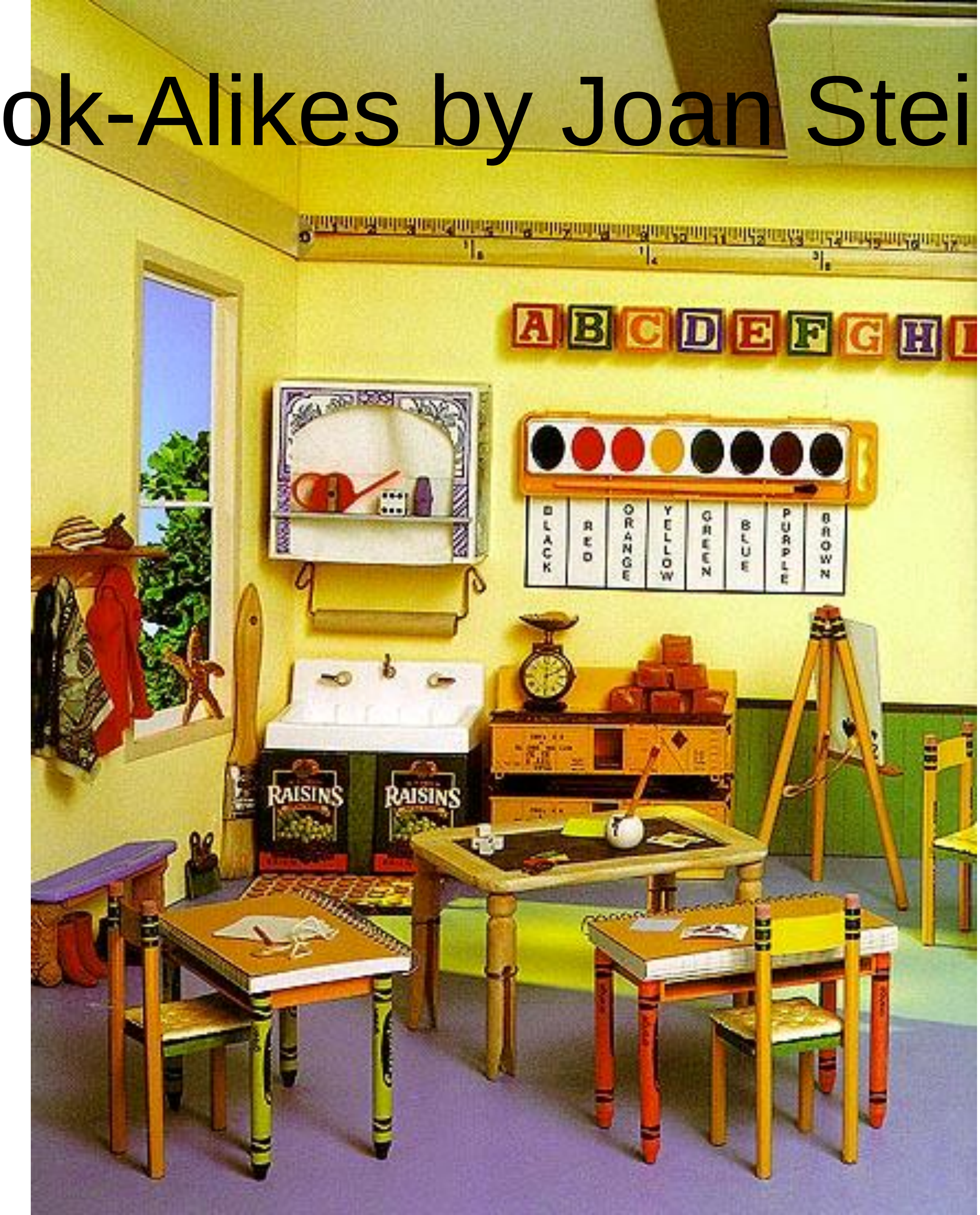


12
A 13 C
14

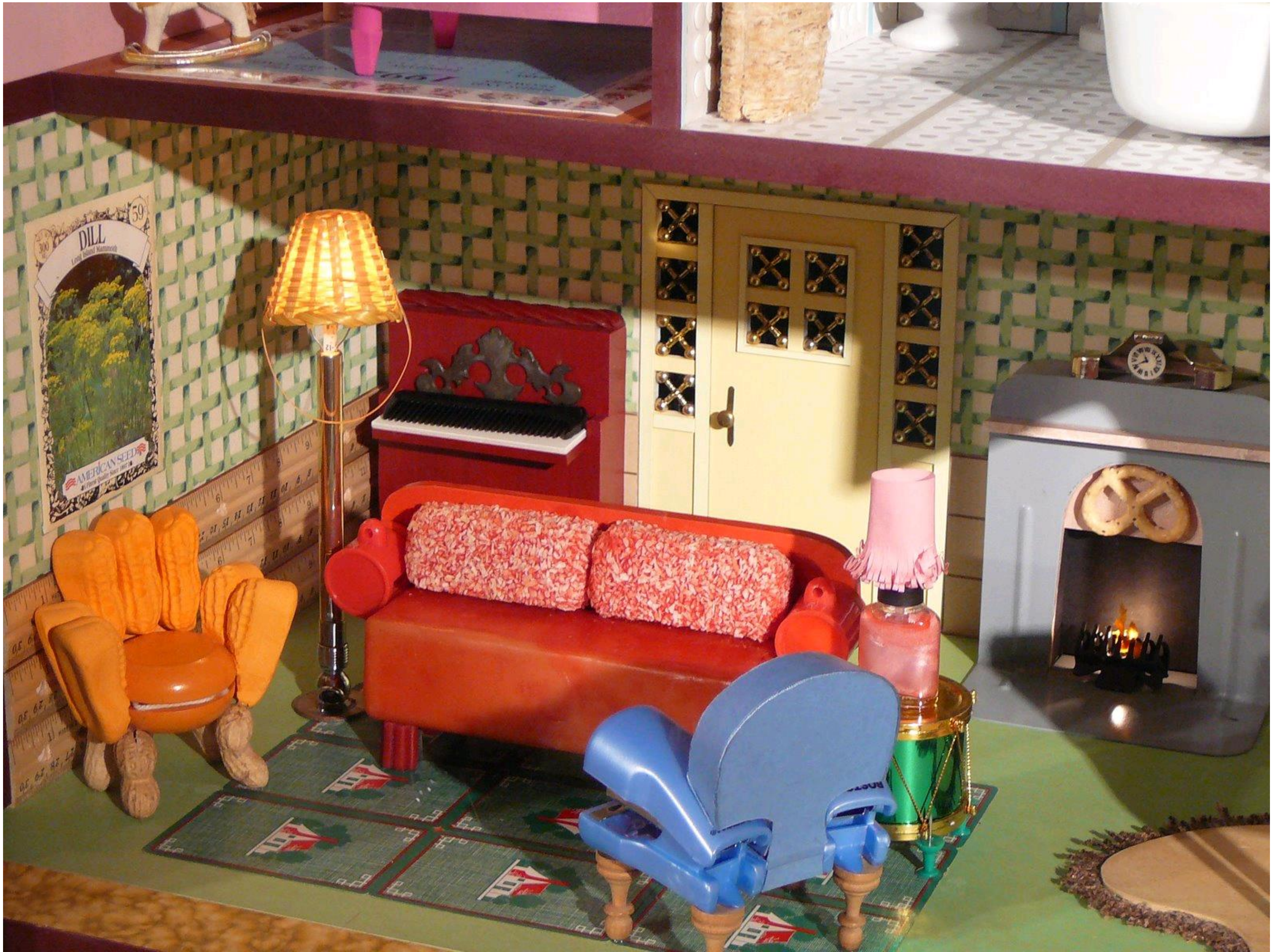
Look-Alikes by Joan Steiner



Look-Alikes by Joan Steiner

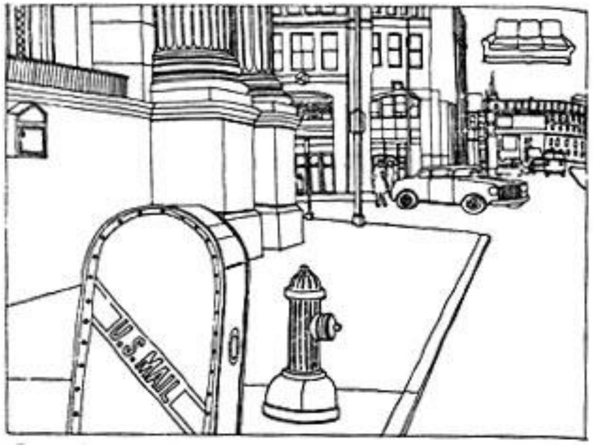


Look-Alikes by Joan Steiner



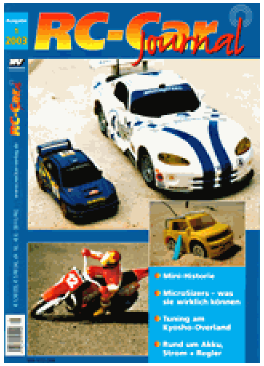
Biederman 1982

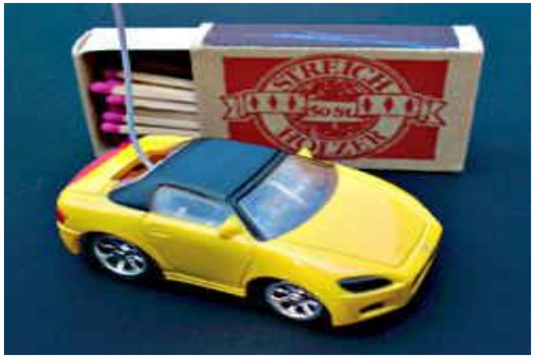
- Pictures shown for 150 ms.
- Objects in appropriate context were detected more accurately than objects in an inappropriate context.
- Scene consistency affects object detection.

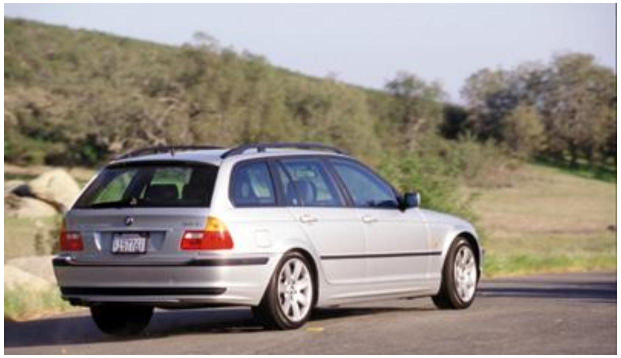


Why is context important?

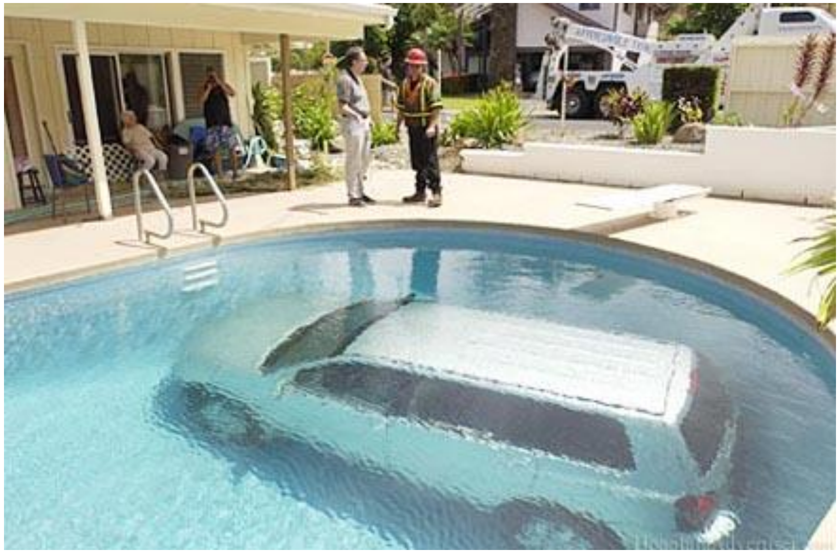
- Changes the interpretation of an object (or its function)

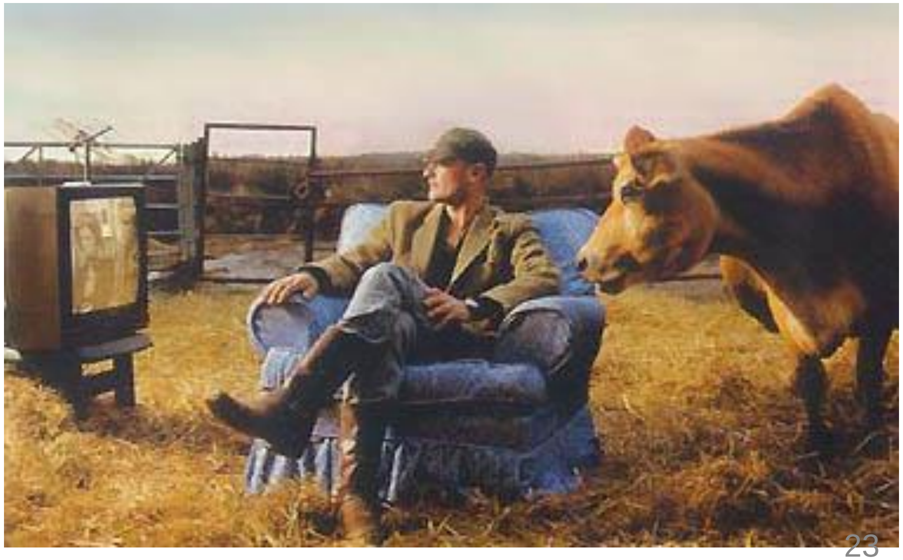






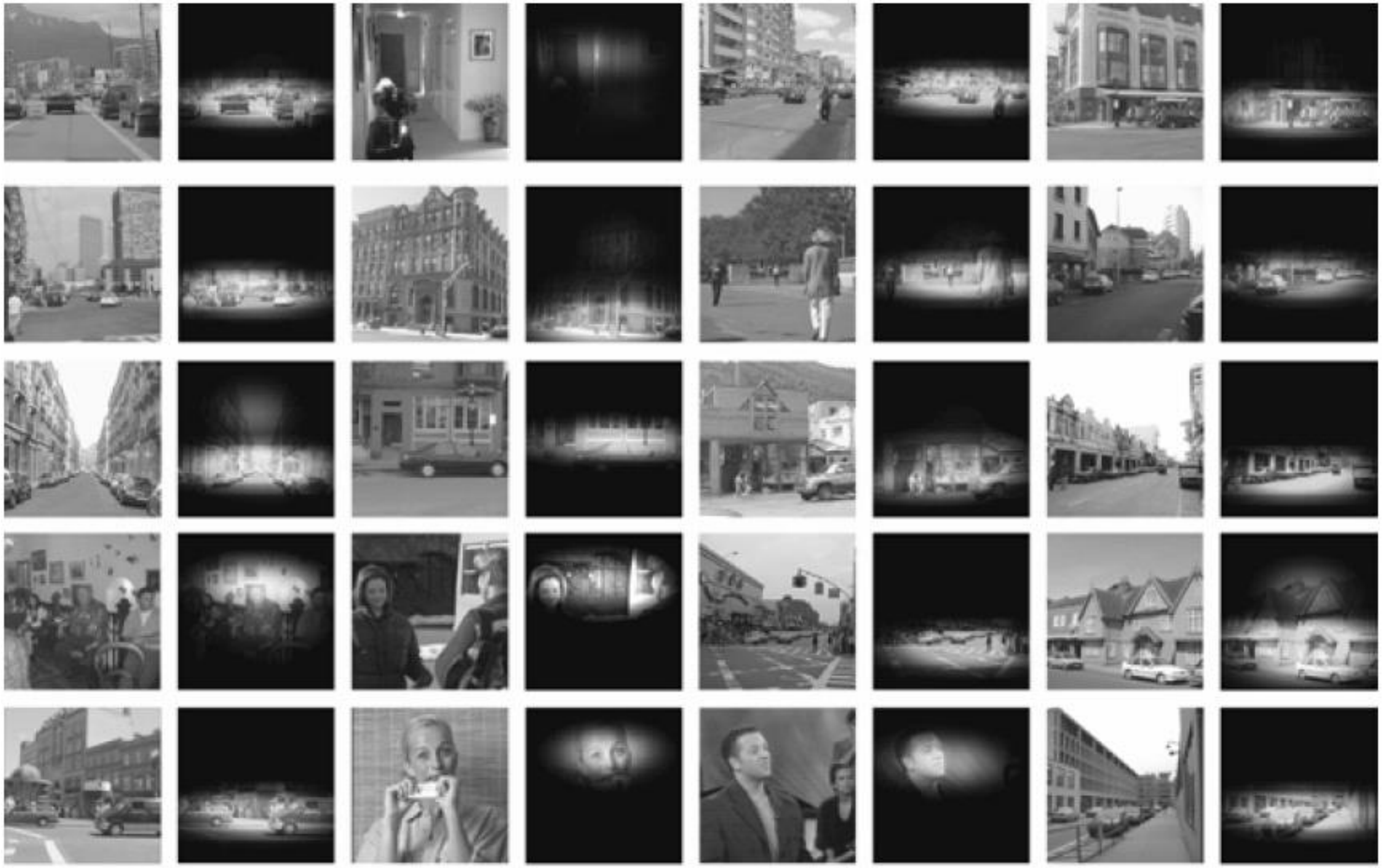
- Context defines what an unexpected event is





The Context Challenge

- http://web.mit.edu/torralba/www/carsAndFacesInContext.html



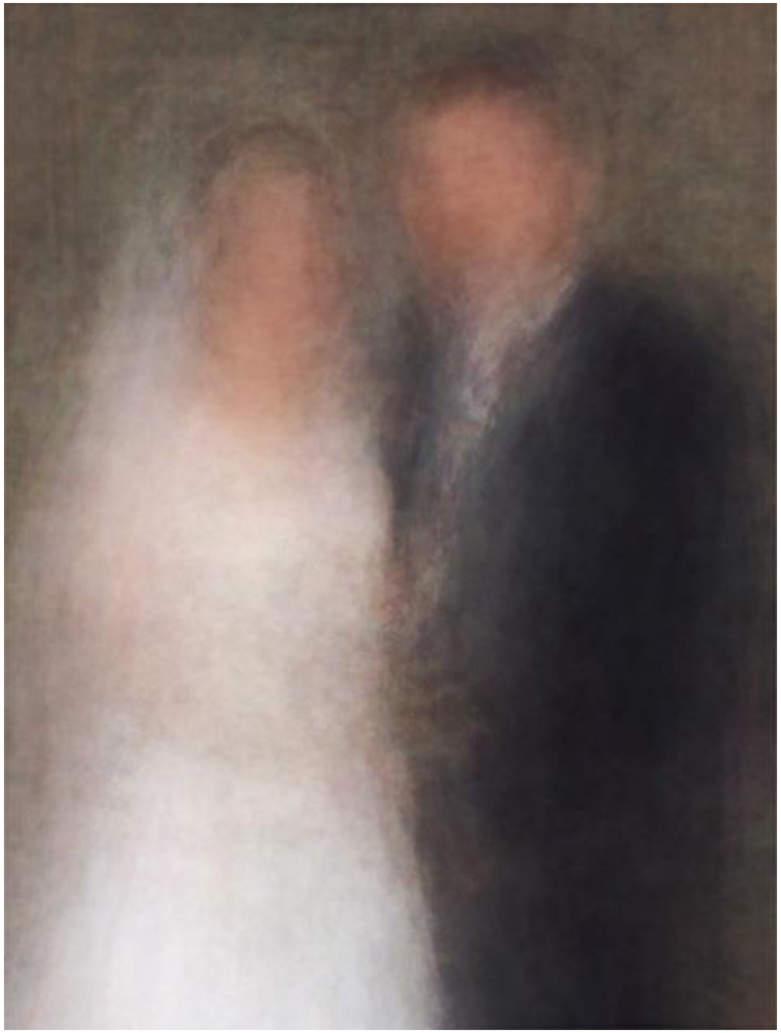
No local face detector! Just context from Scene Statistics

There are many types of context

- **Local pixels**
  - window, surround, image neighborhood, object boundary/shape, global image statistics
- **2D Scene Gist**
  - global image statistics
- **3D Geometric**
  - 3D scene layout, support surface, surface orientations, occlusions, contact points, etc.
- **Semantic**
  - event/activity depicted, scene category, objects present in the scene and their spatial extents, keywords
- **Photogrammetric**
  - camera height orientation, focal length, lens distortion, radiometric, response function
- **Illumination**
  - sun direction, sky color, cloud cover, shadow contrast, etc.
- **Geographic**
  - GPS location, terrain type, land use category, elevation, population density, etc.
- **Temporal**
  - nearby frames of video, photos taken at similar times, videos of similar scenes, time of capture
- **Cultural**
  - photographer bias, dataset selection bias, visual cliches, etc.

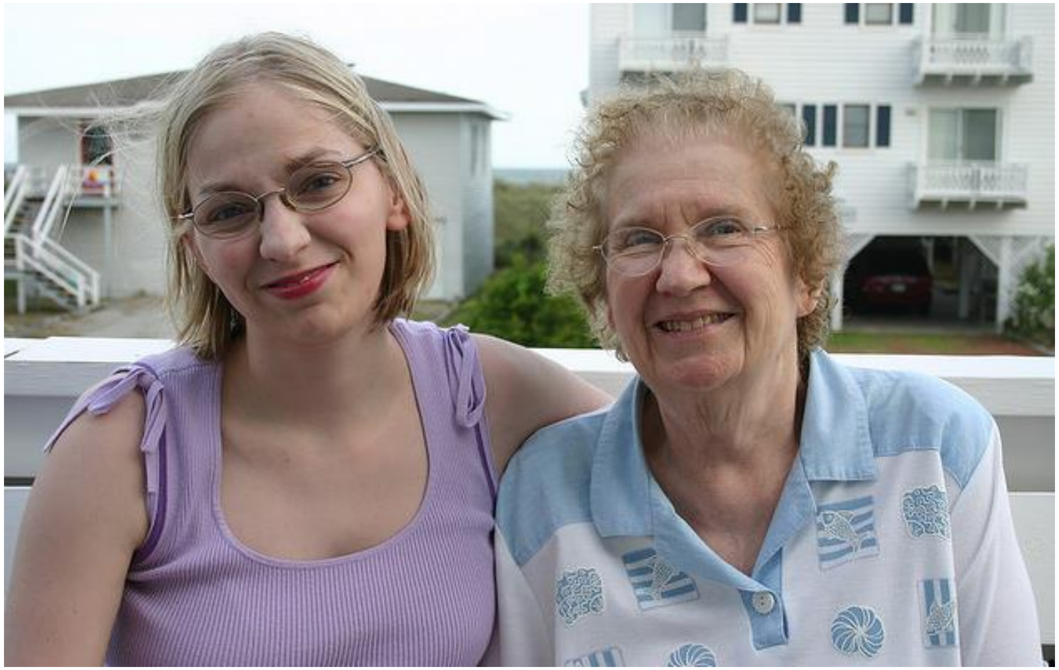
from Divvala et al. CVPR 2009

Cultural context



Jason Salavon: http://salavon.com/SpecialMoments/Newlyweds.shtml

Cultural context



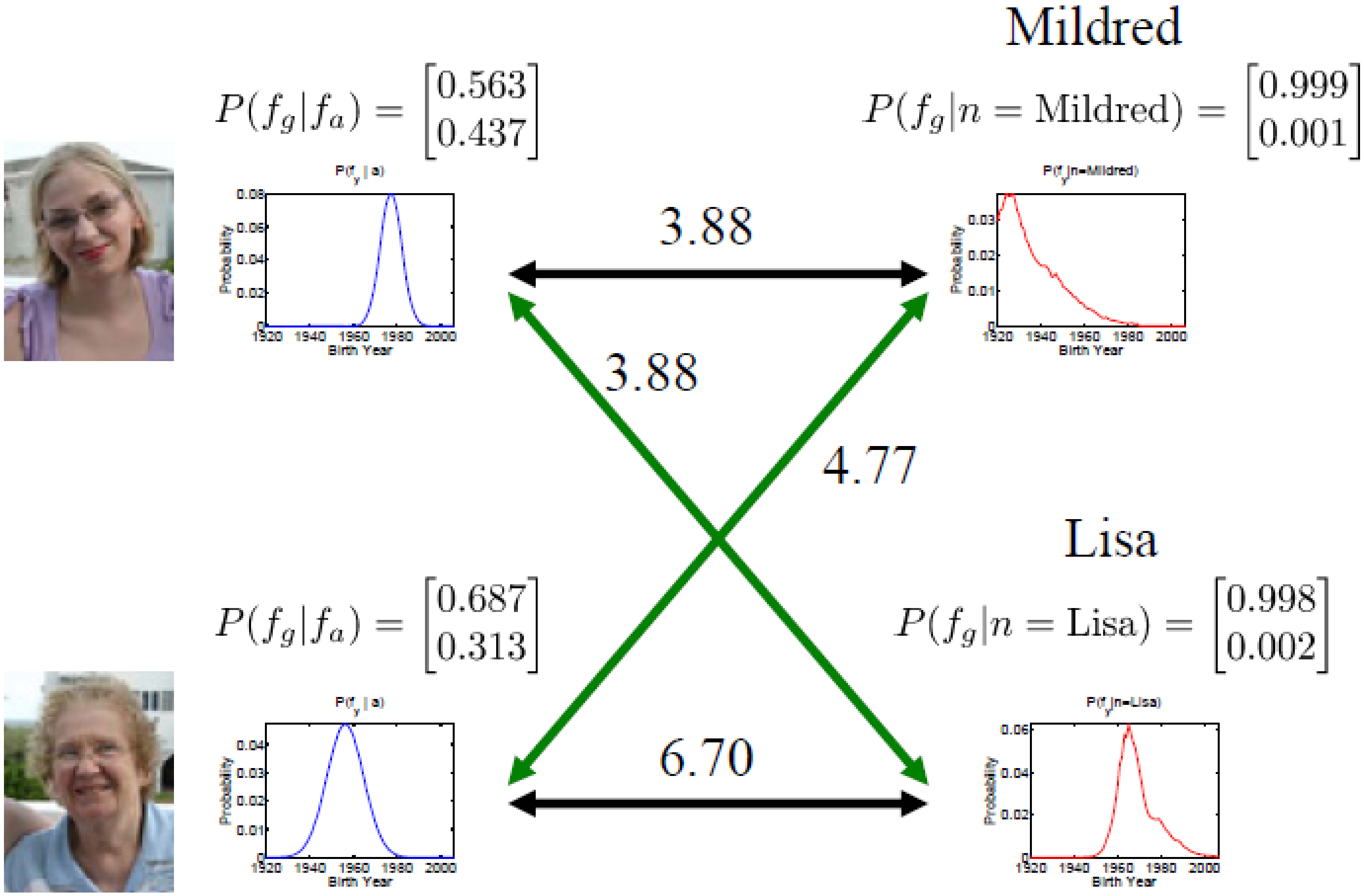
Who is Mildred? Who is Lisa?

Andrew Gallagher: http://chenlab.ece.cornell.edu/people/Andy/projectpage_names.html

Cultural context

Age given Appearance

Age given Name



Mildred

$$P(f_g|f_a) = \begin{bmatrix}0.563\\0.437\end{bmatrix}$$

$$P(f_g|n = \text{Mildred}) = \begin{bmatrix}0.999\\0.001\end{bmatrix}$$

Lisa

$$P(f_g|f_a) = \begin{bmatrix}0.687\\0.313\end{bmatrix}$$

$$P(f_g|n = \text{Lisa}) = \begin{bmatrix}0.998\\0.002\end{bmatrix}$$

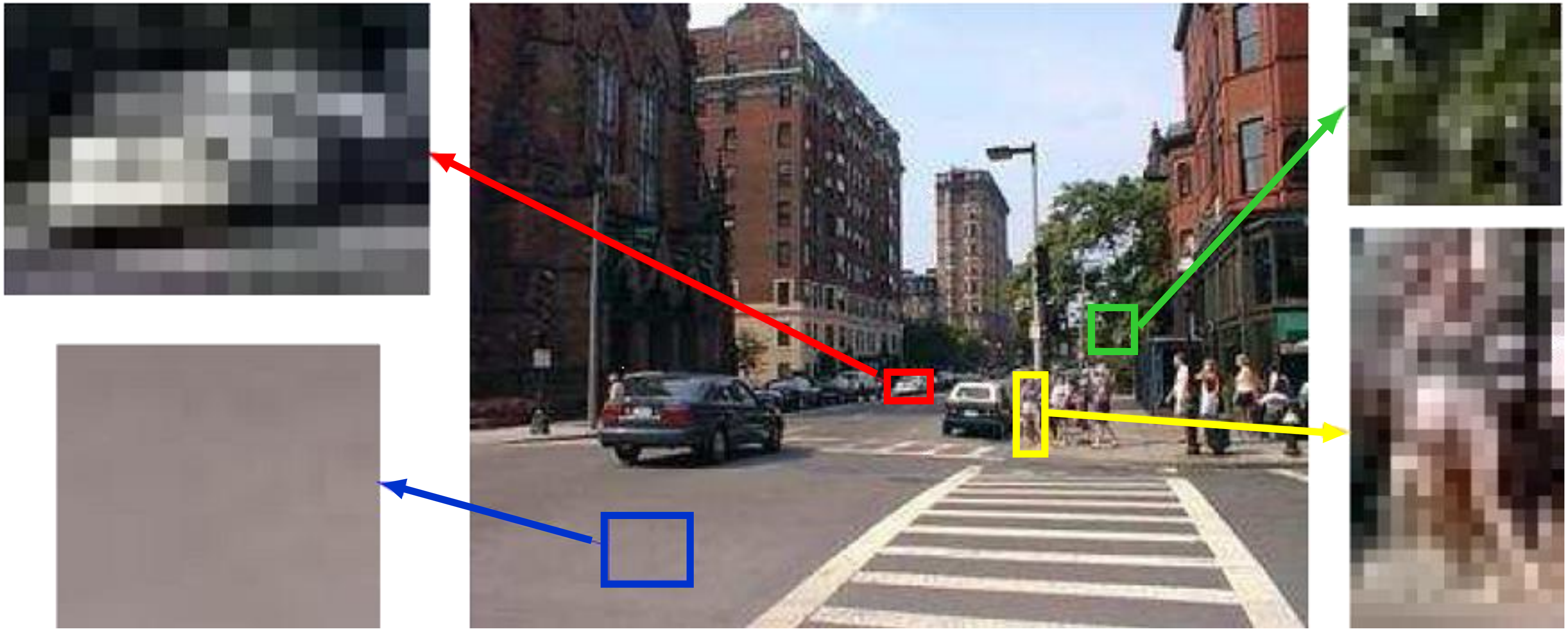
Andrew Gallagher: http://chenlab.ece.cornell.edu/people/Andy/projectpage_names.html

Spatial layout is especially important

1. Context for recognition

Spatial layout is especially important

1. Context for recognition



Slide: Derek Hoiem

Spatial layout is especially important

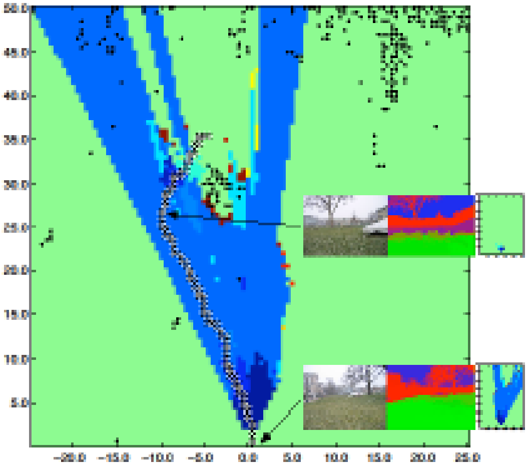
1. Context for recognition
2. Scene understanding

Slide: Derek Hoiem

Spatial layout is especially important

1. Context for recognition
2. Scene understanding
3. Many direct applications
   a) Assisted driving
   b) Robot navigation/interaction
   c) 2D to 3D conversion for 3D TV
   d) Object insertion

3D Reconstruction: Input, Mesh, Novel View

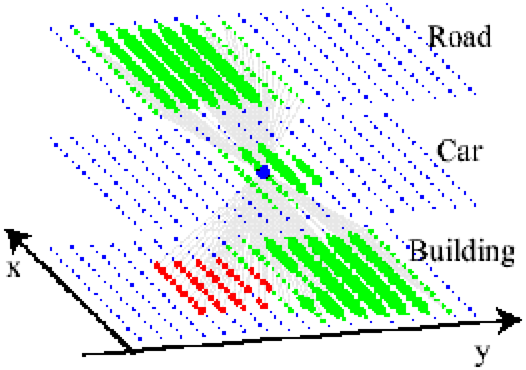


Robot Navigation: Path Planning

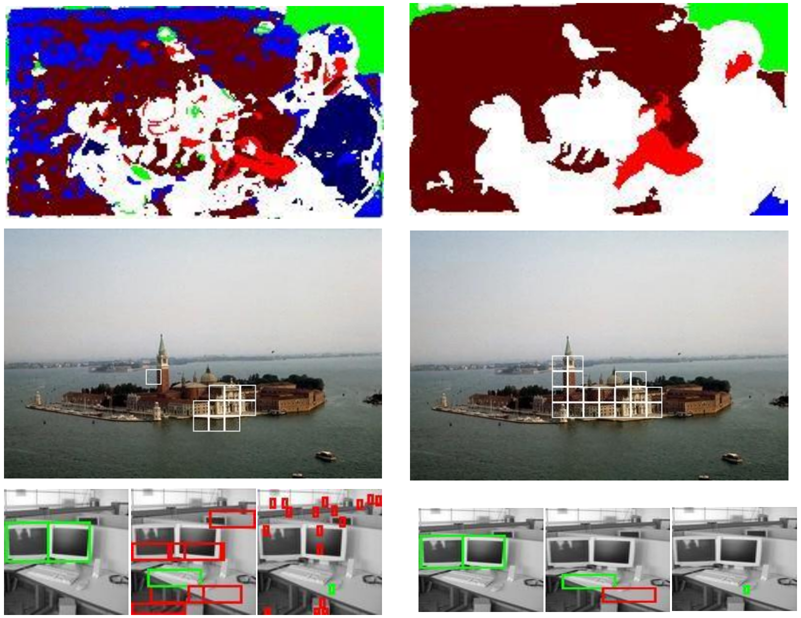
Spatial Layout: 2D vs. 3D

Slide: Derek Hoiem

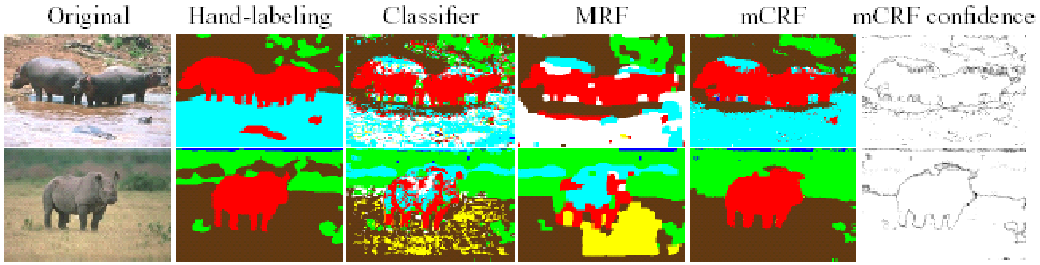
Context in Image Space



[Torralba Murphy Freeman 2004]



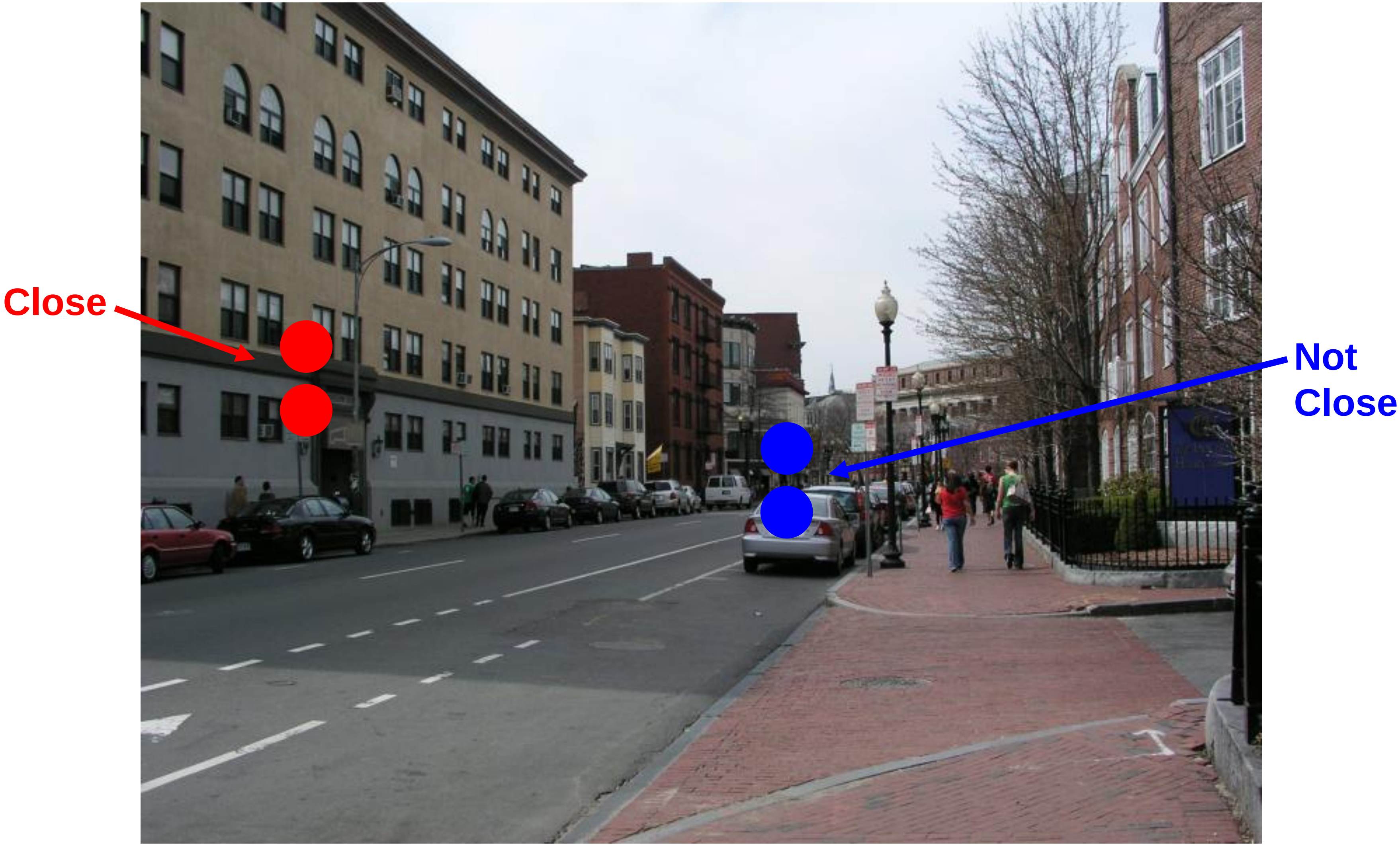
[Kumar Hebert 2005]



[He Zemel Cerreira-Perpiñán 2004]

Slide: Derek Hoiem

But object relations are in 3D…



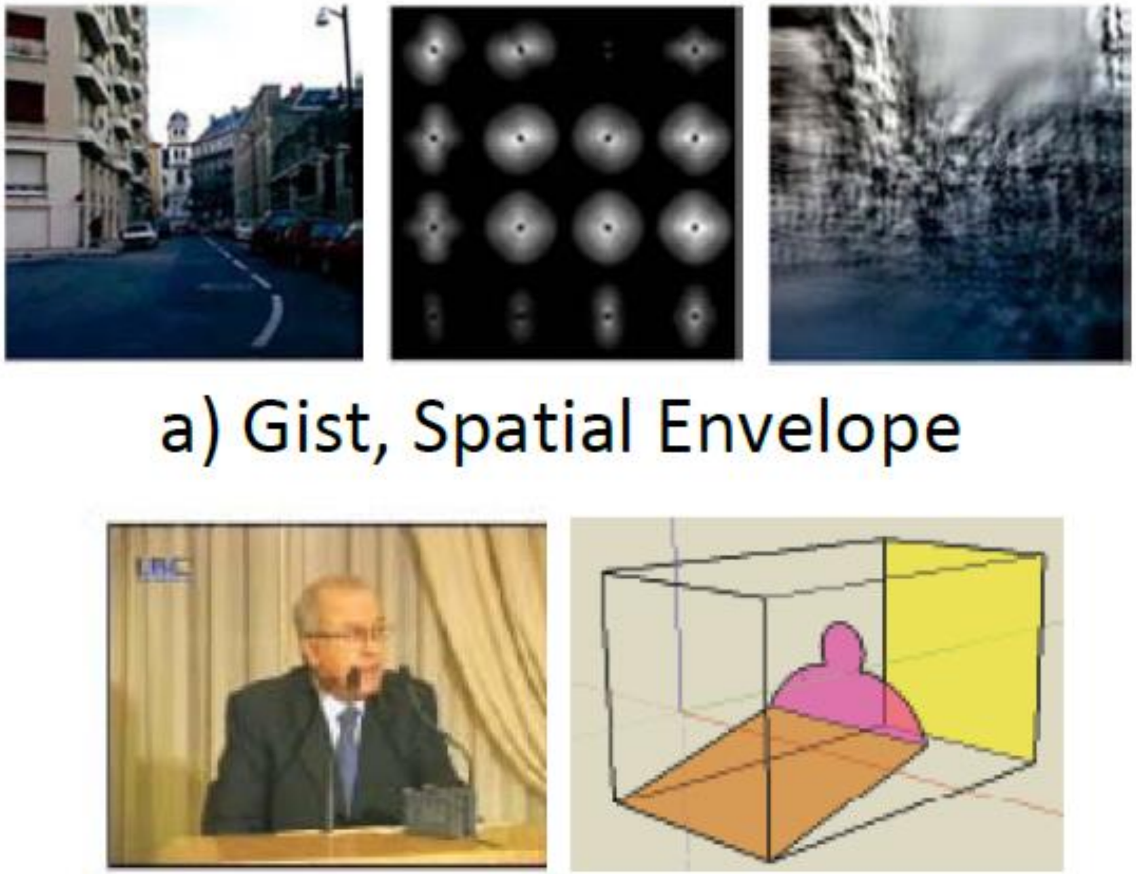
Slide: Derek Hoiem

How to represent scene space?

Slide: Derek Hoiem

Wide variety of possible representations

Scene-Level Geometric Description

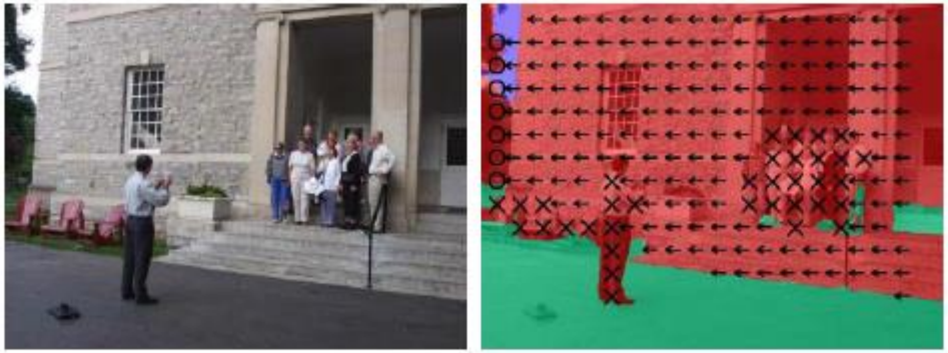


a) Gist, Spatial Envelope

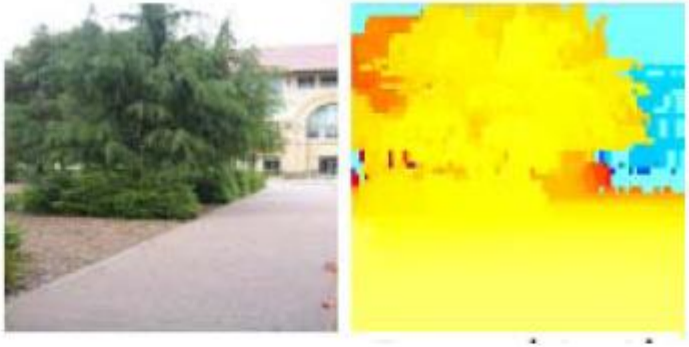
b) Stages

Figs from Hoiem/Savarese Draft

Retinotopic Maps



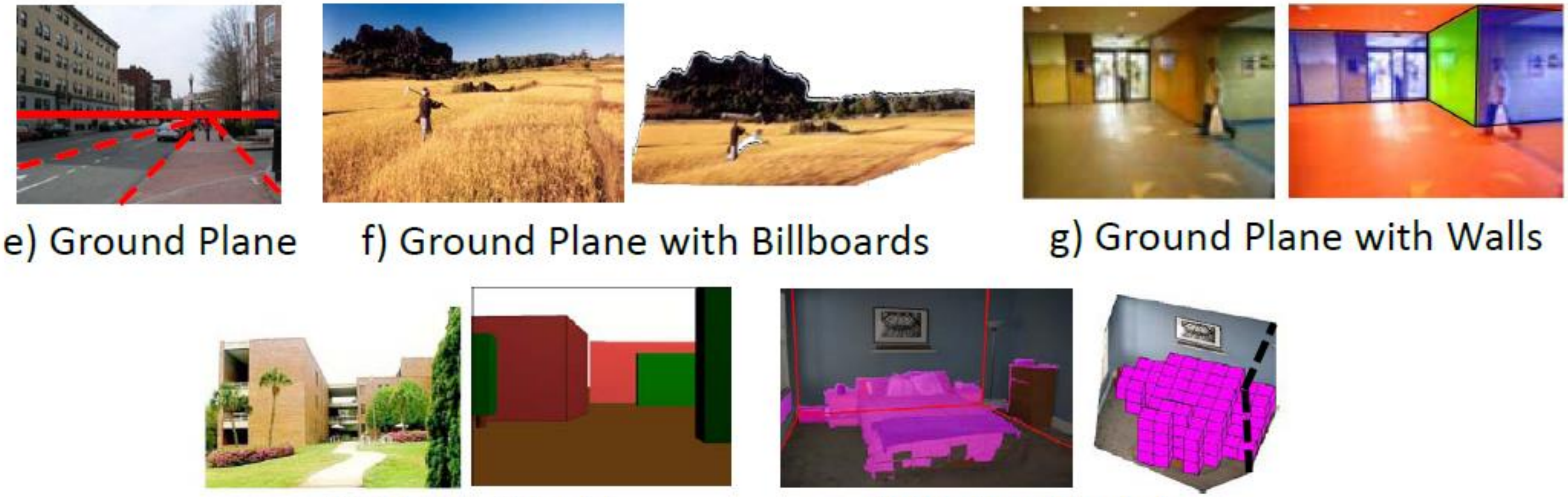
c) Geometric Context



d) Depth Maps

Figs from Hoiem/Savarese Draft

Highly Structured 3D Models



e) Ground Plane

f) Ground Plane with Billboards

g) Ground Plane with Walls

h) Blocks World

i) 3D Box Model

Figs from Hoiem/Savarese Draft

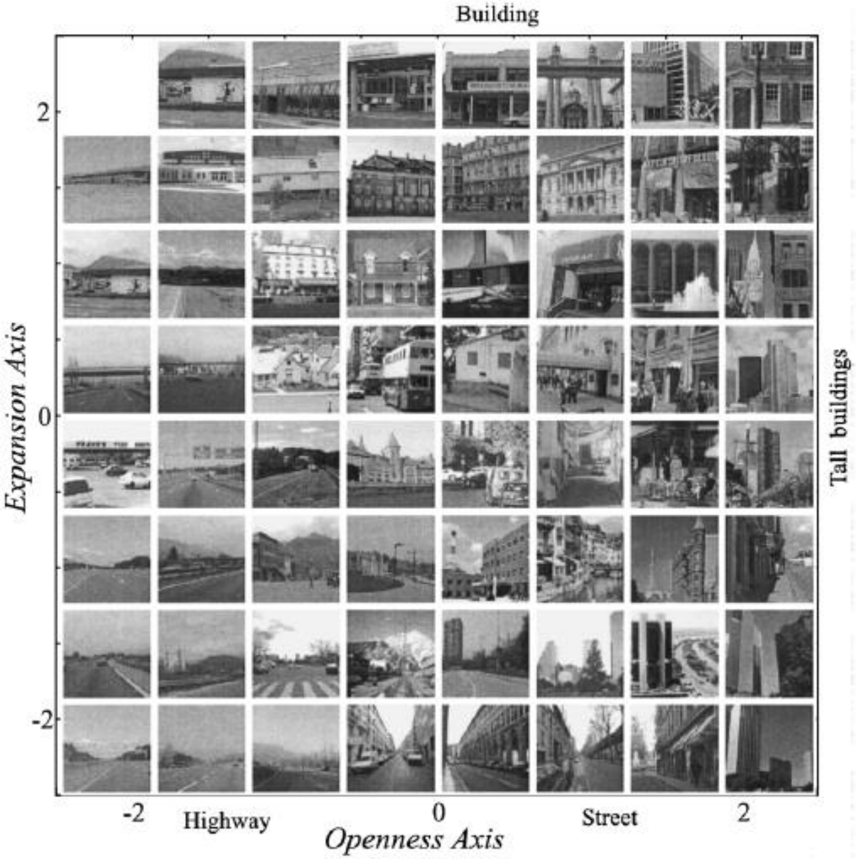
Key Trade-offs

- Level of detail: rough “gist”, or detailed point cloud?
  - Precision vs. accuracy
  - Difficulty of inference
- Abstraction: depth at each pixel, or ground planes and walls?
  - What is it for: e.g., metric reconstruction vs. navigation

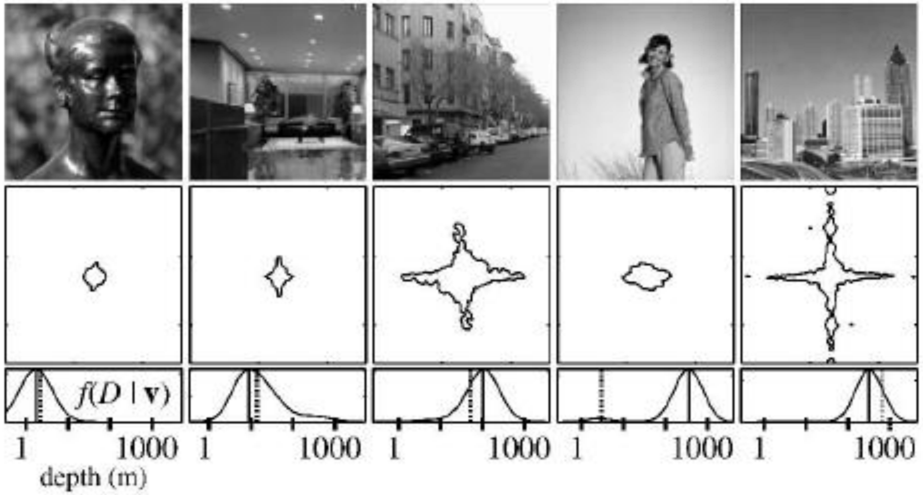
Slide: Derek Hoiem

Low detail, Low abstraction

Holistic Scene Space: “Gist”



Oliva & Torralba 2001

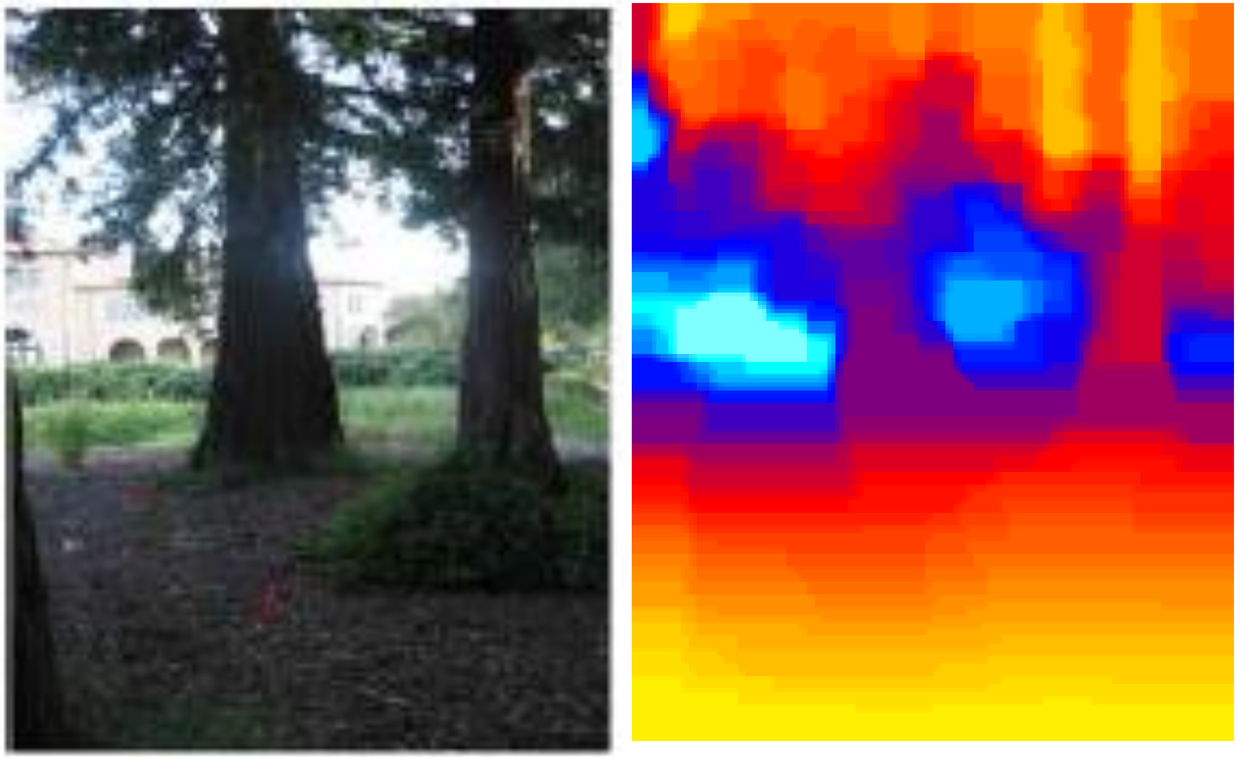


Torralba & Oliva 2002

Slide: Derek Hoiem

High detail, Low abstraction

Depth Map

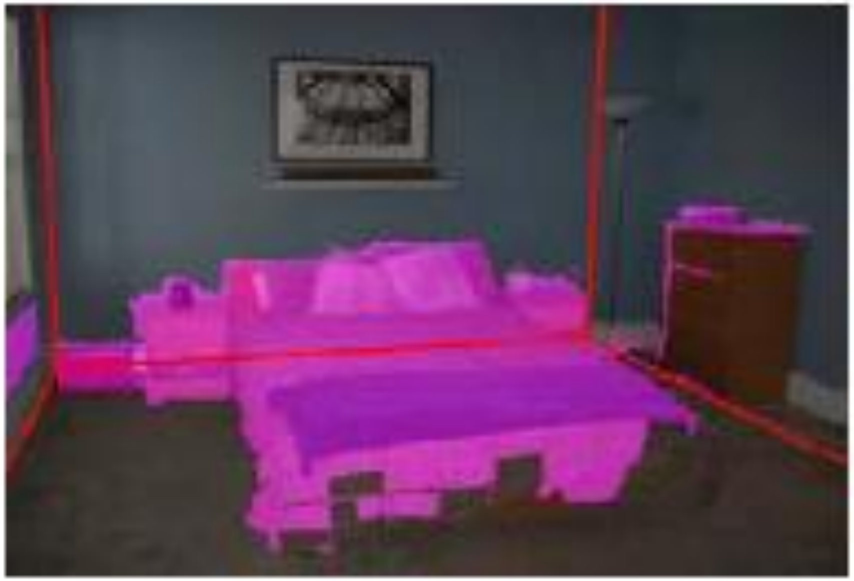


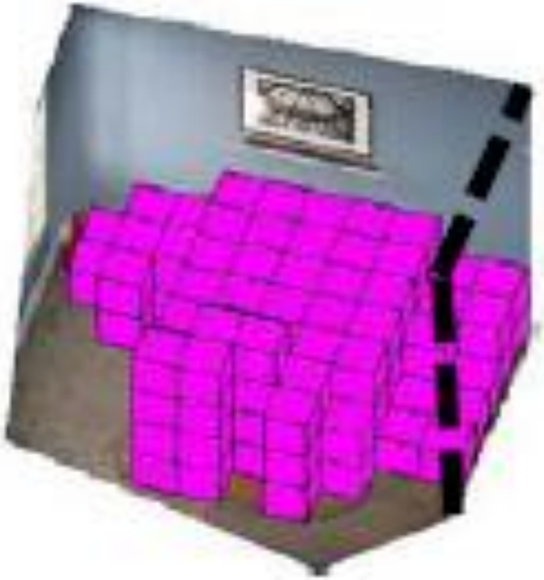
Saxena, Chung & Ng 2005, 2007

Slide: Derek Hoiem

Medium detail, High abstraction

Room as a Box

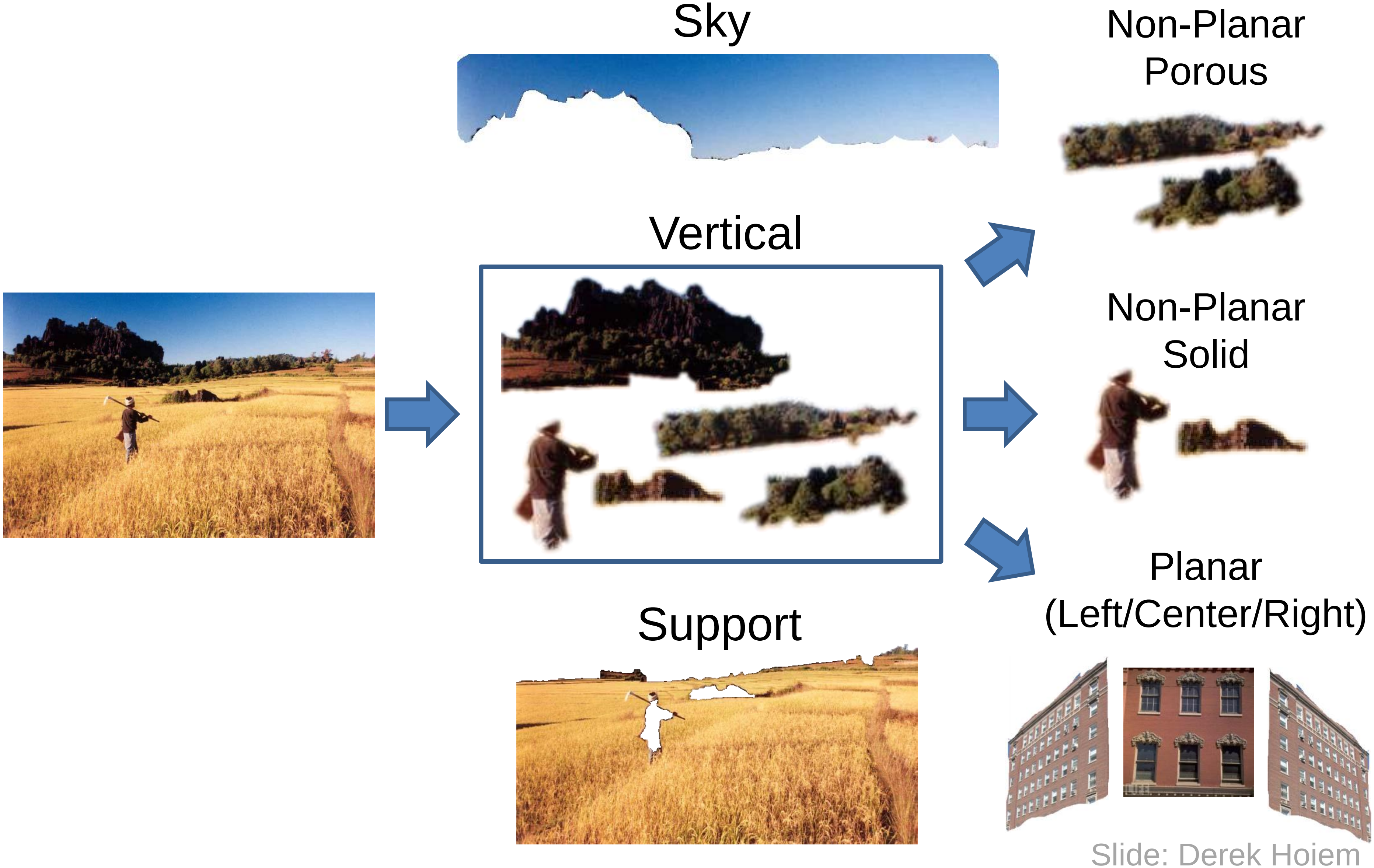




Hedau Hoiem Forsyth 2009

Slide: Derek Hoiem

Surface Layout: describe 3D surfaces with geometric classes



Sky

Vertical

Support

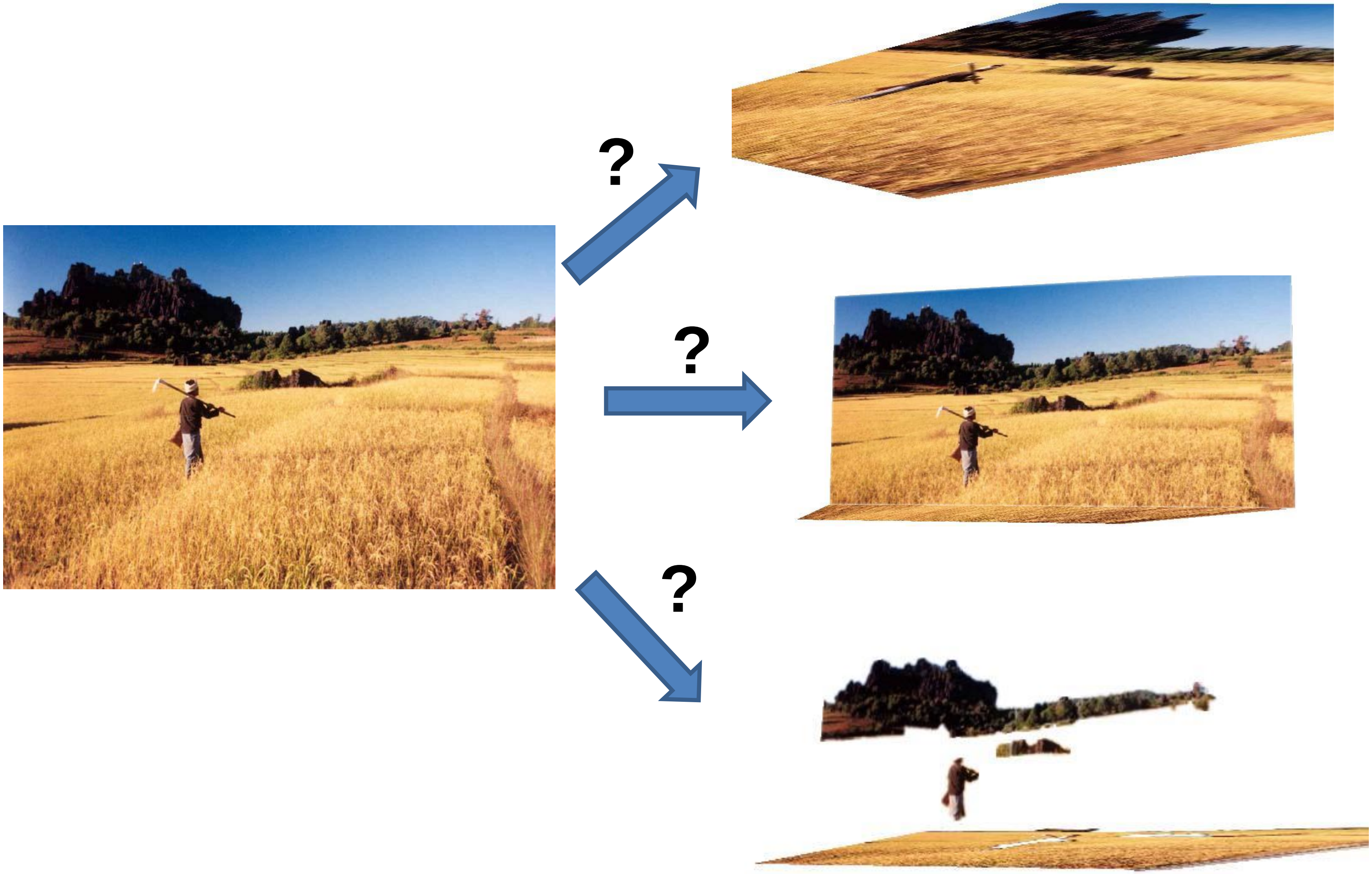
Non-Planar Porous

Non-Planar Solid

Planar (Left/Center/Right)

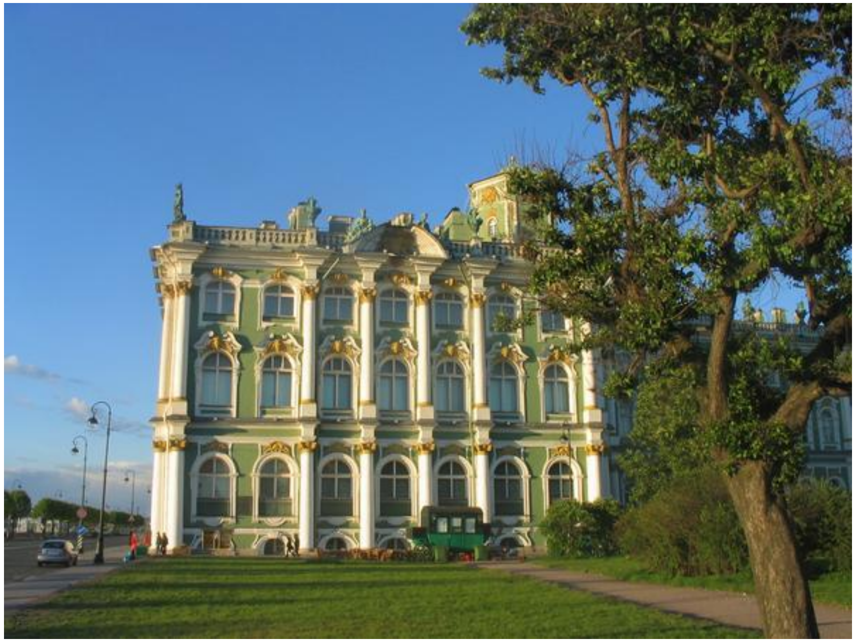
Slide: Derek Hoiem

The challenge



Our World is Structured

Abstract World



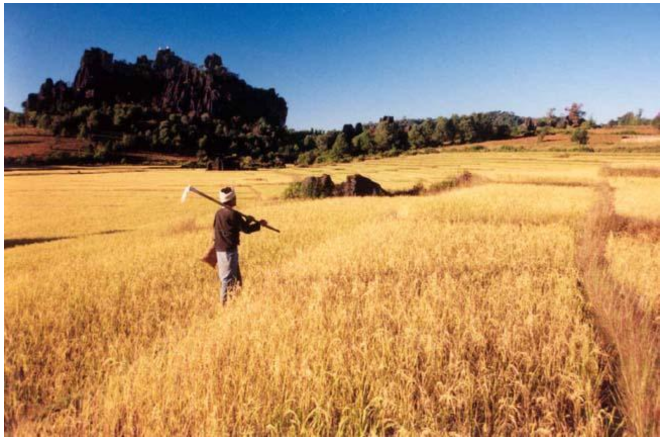
Our World

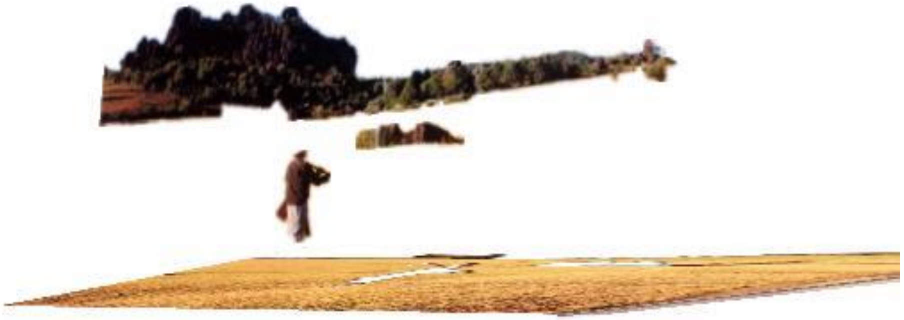
Image Credit (left): F. Cunin
and M.J. Sailor, UCSD

Learn the Structure of the World

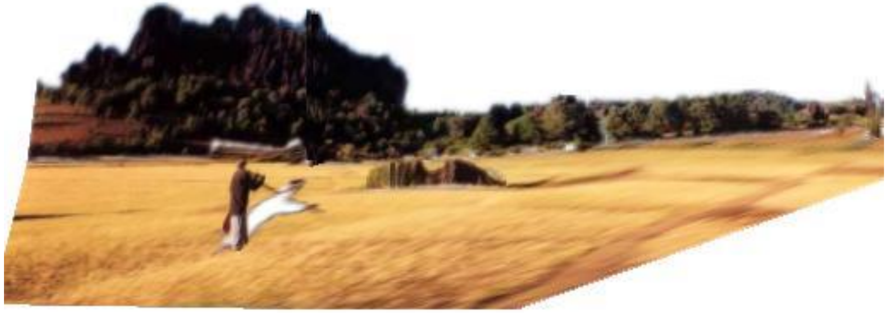
Training Images

Infer the most likely interpretation



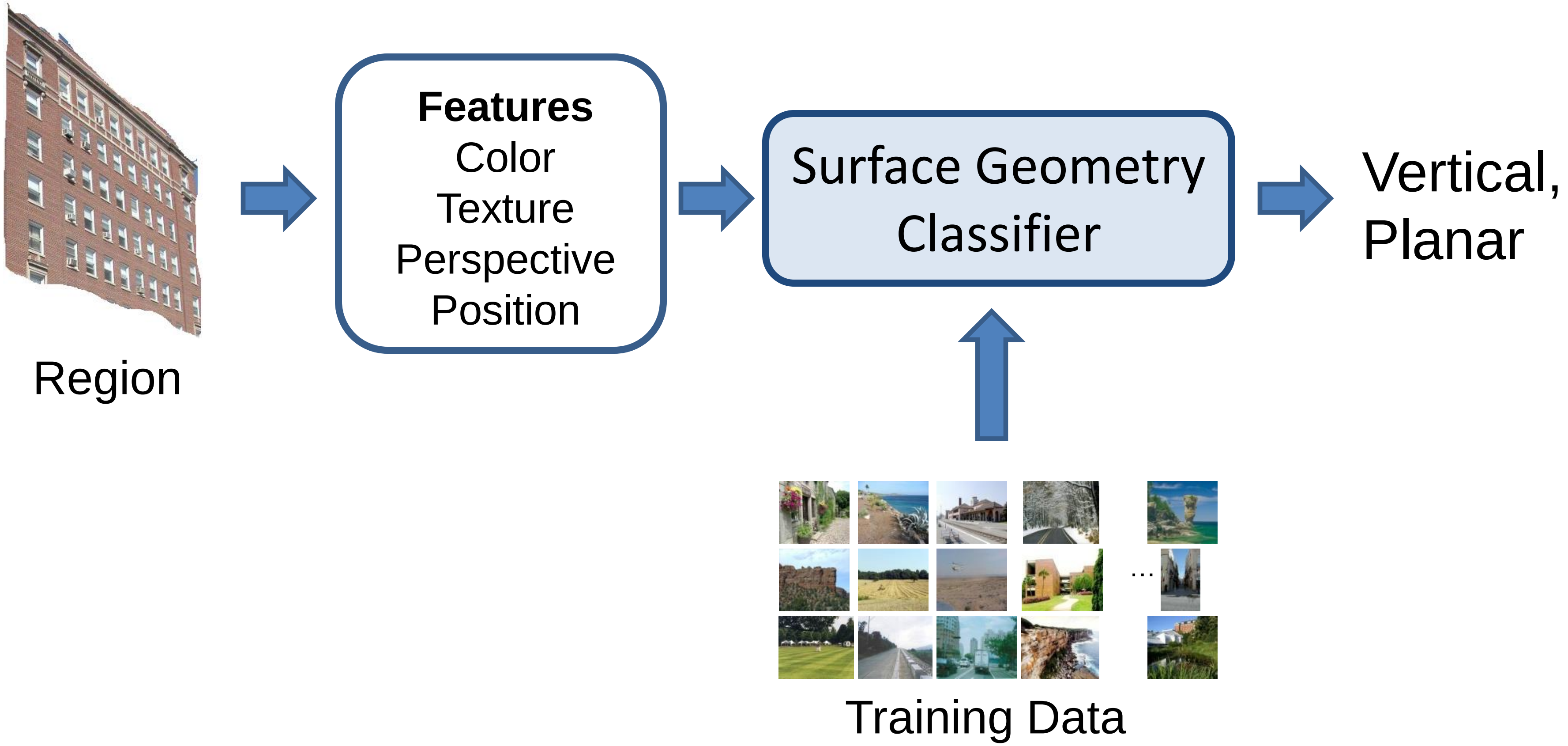


Unlikely



Likely

Geometry estimation as recognition



Region

Features
Color
Texture
Perspective
Position

Surface Geometry Classifier

Vertical, Planar

Training Data

...

Slide: Derek Hoiem

Use a variety of image cues

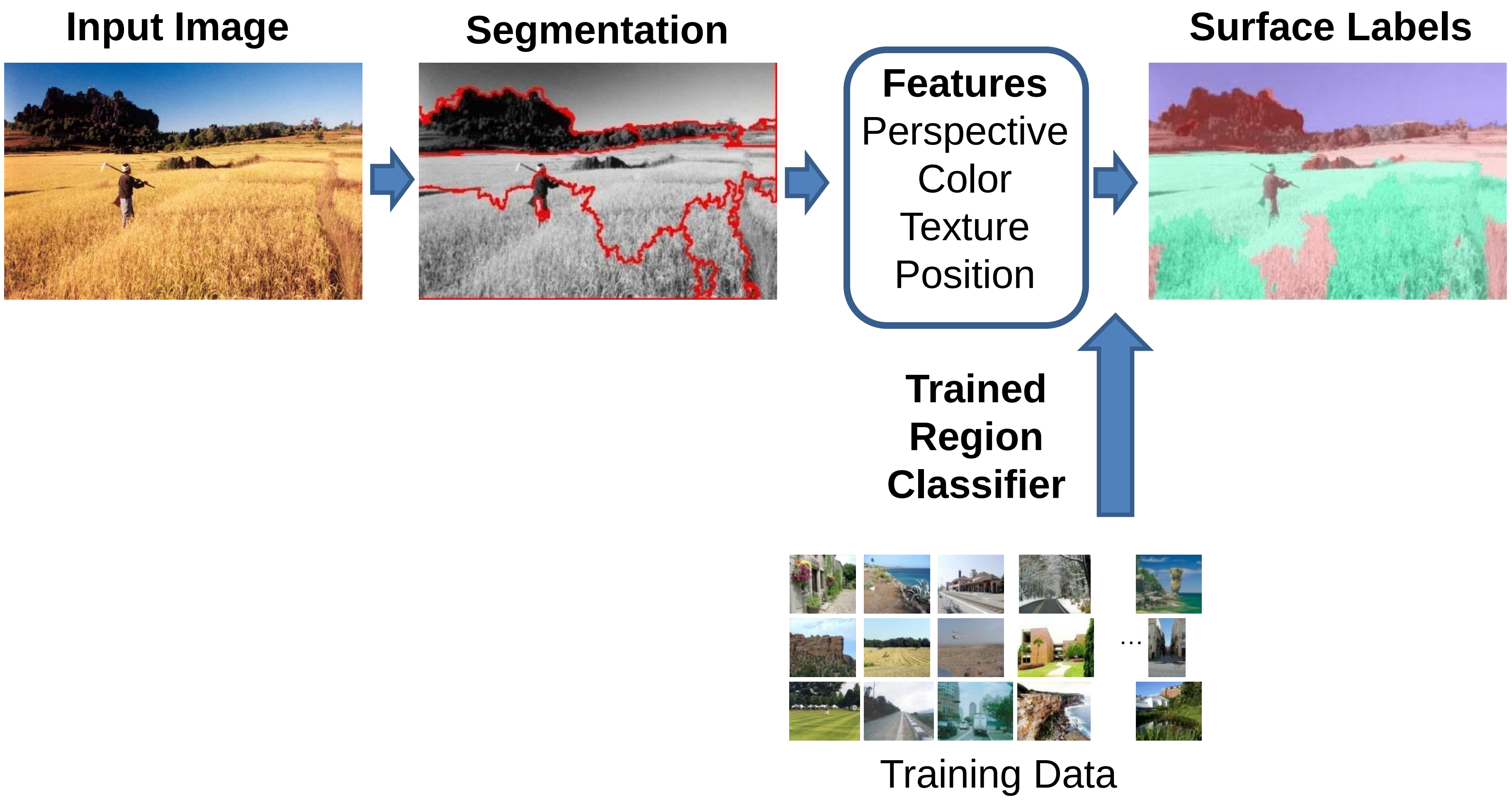
Vanishing points, lines

Color, texture, image location

Texture gradient

Slide: Derek Hoiem

Surface Layout Algorithm



Input Image

Segmentation

Features
Perspective
Color
Texture
Position

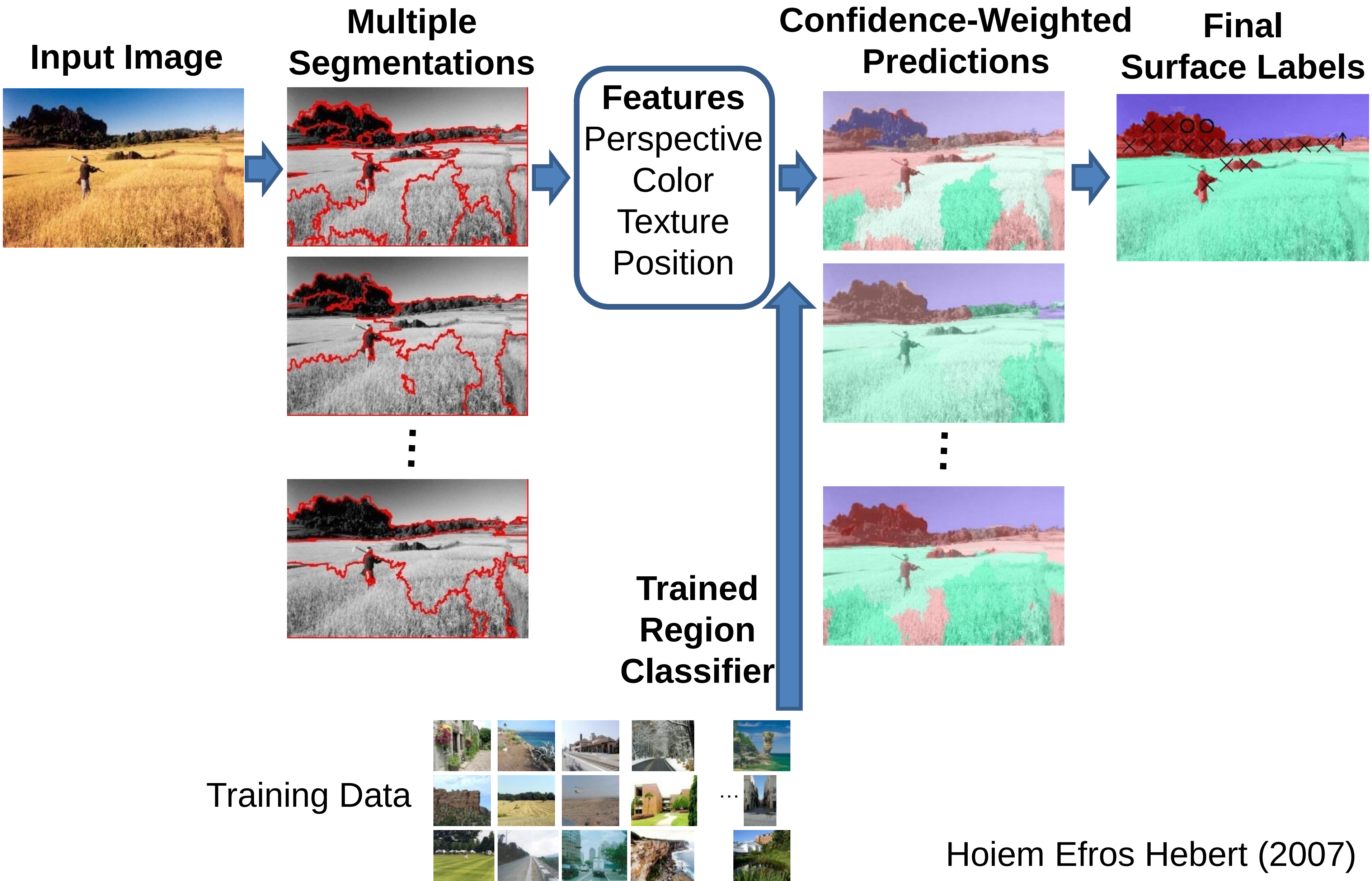
Surface Labels

Trained Region Classifier

Training Data

Hoiem Efros Hebert (2007)

Surface Layout Algorithm



Input Image

Multiple Segmentations

Features
Perspective
Color
Texture
Position

Confidence-Weighted Predictions

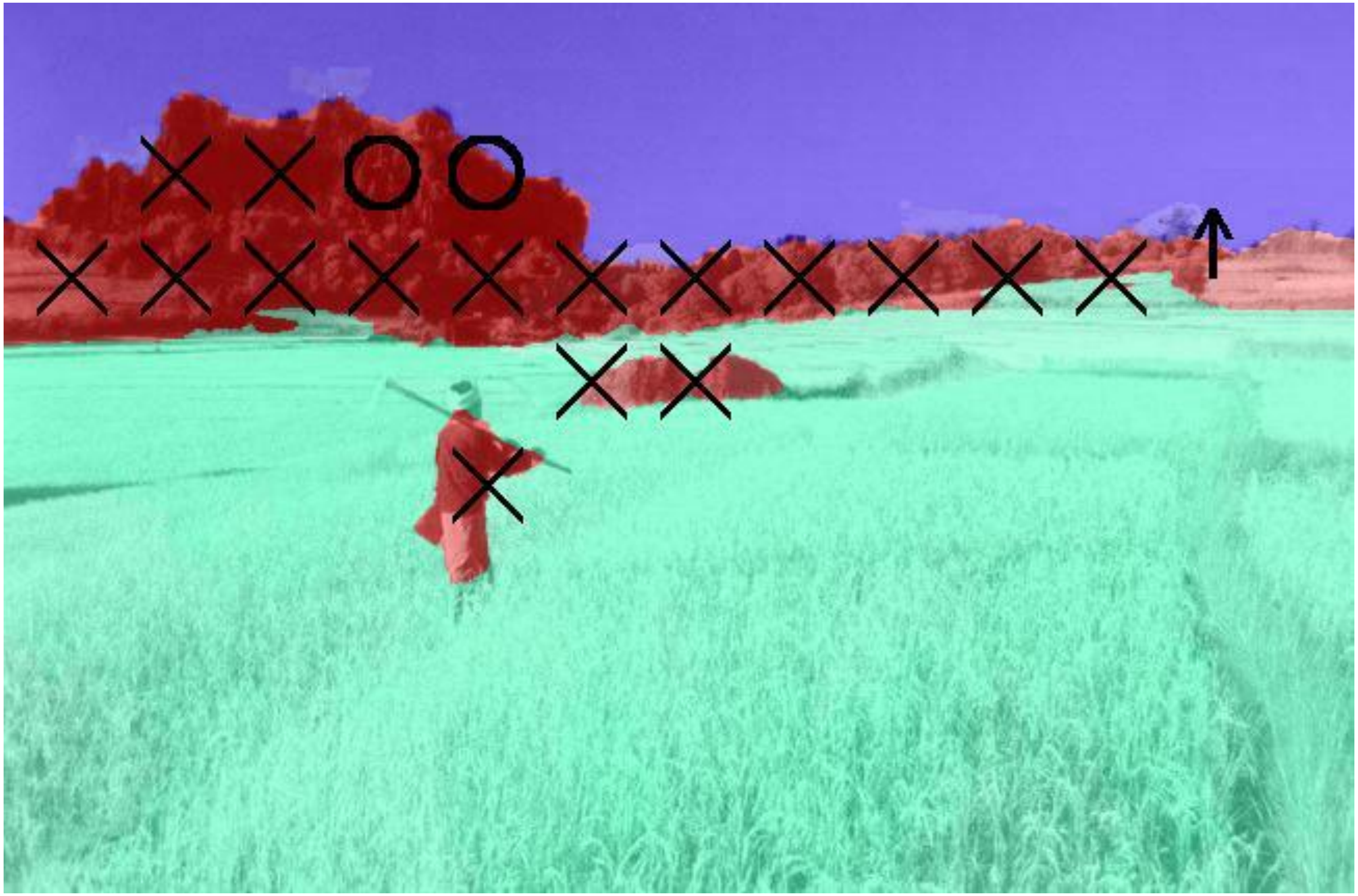
Final Surface Labels

Trained Region Classifier

Training Data

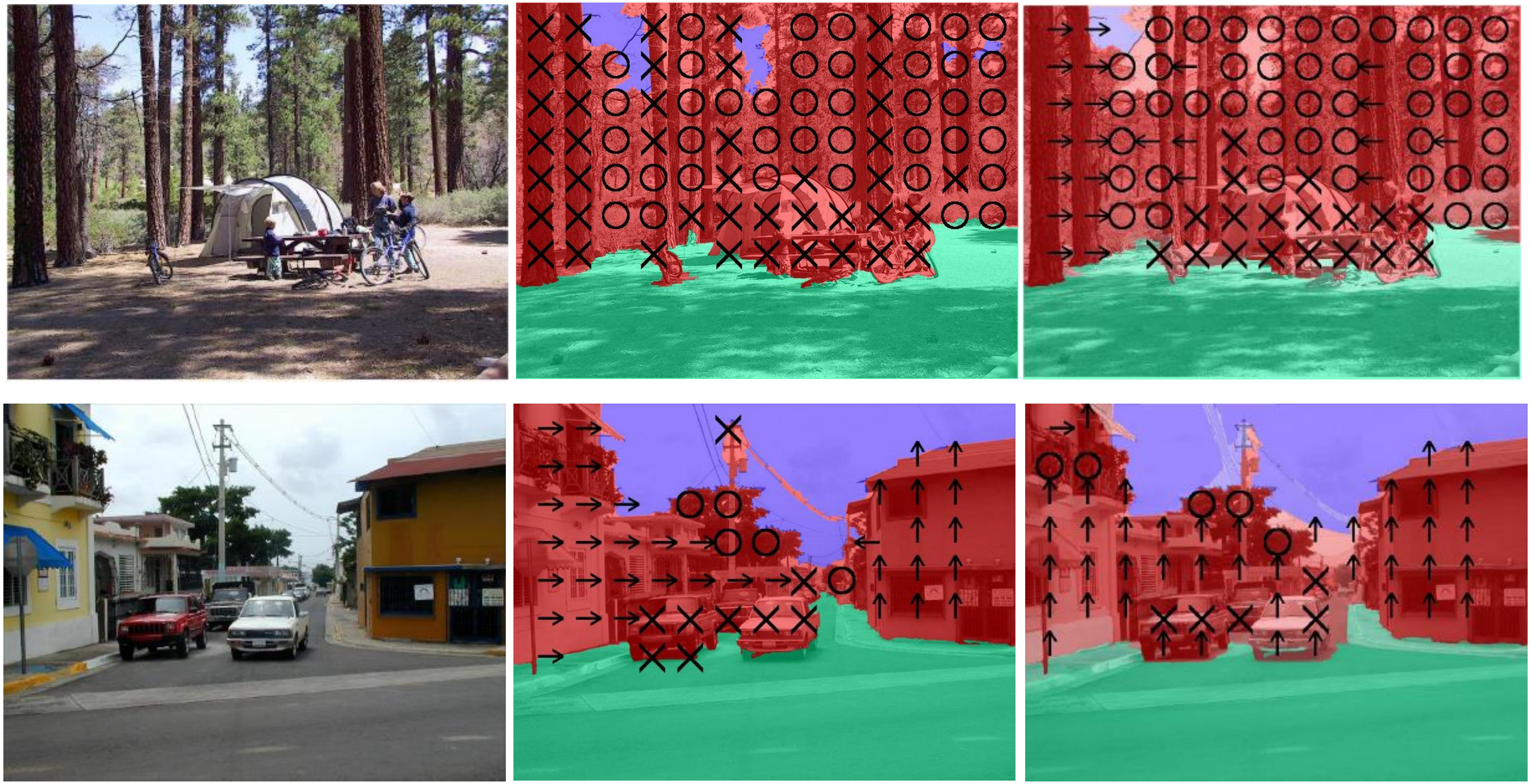
Hoiem Efros Hebert (2007)

Surface Description Result



Slide: Derek Hoiem

Results

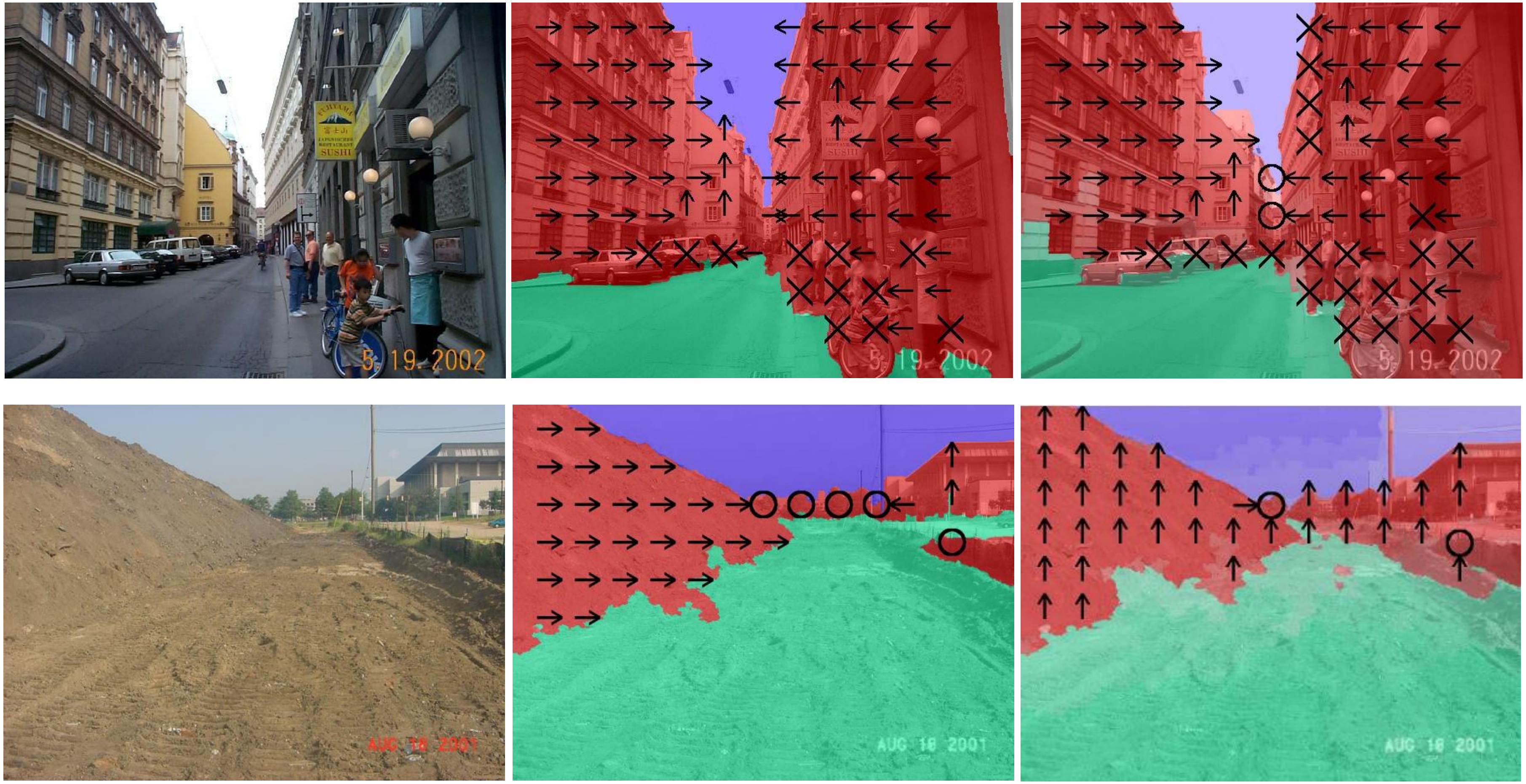


Input Image

Ground Truth

Our Result

Results

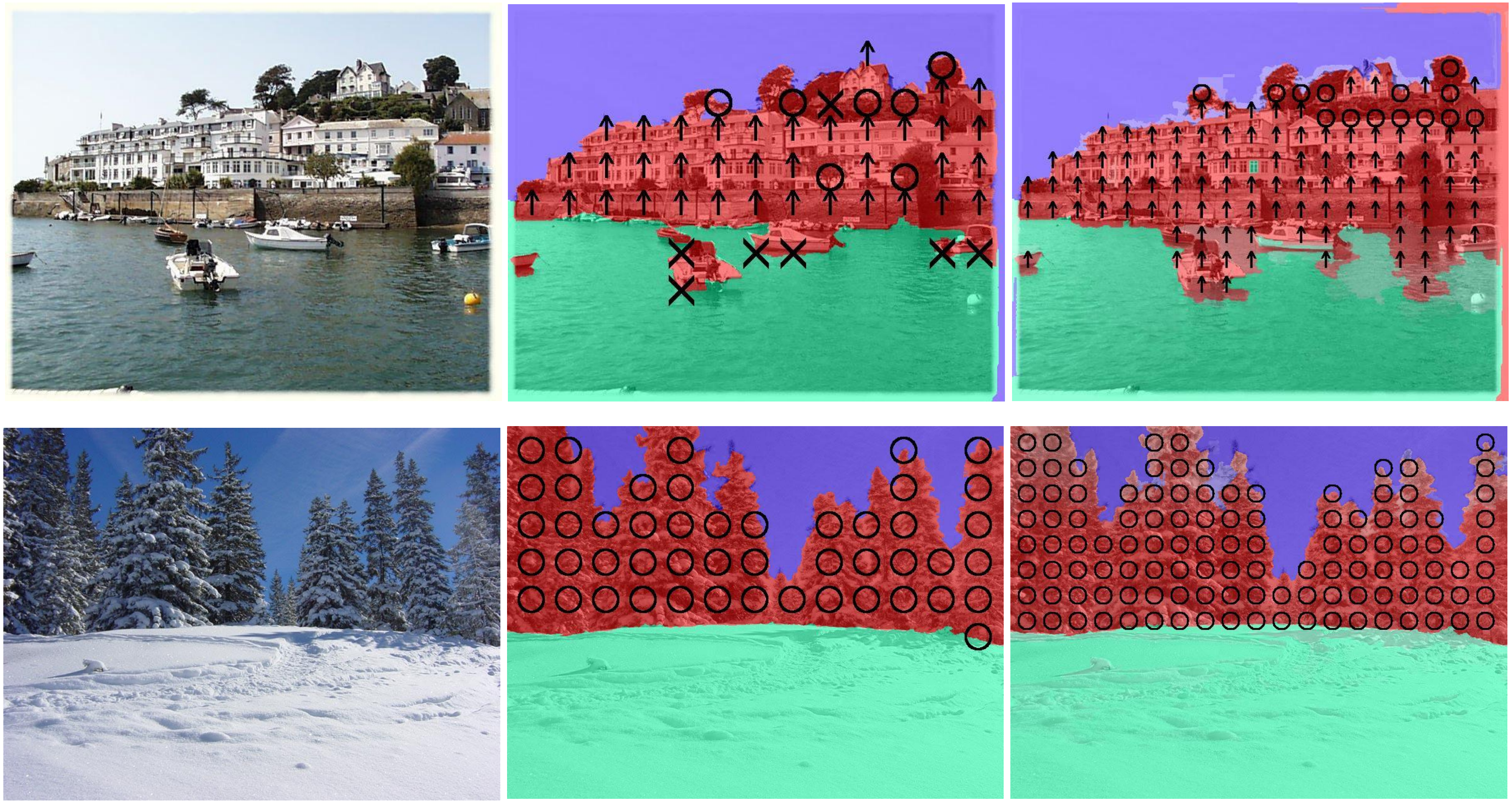


Input Image

Ground Truth

Our Result

Results

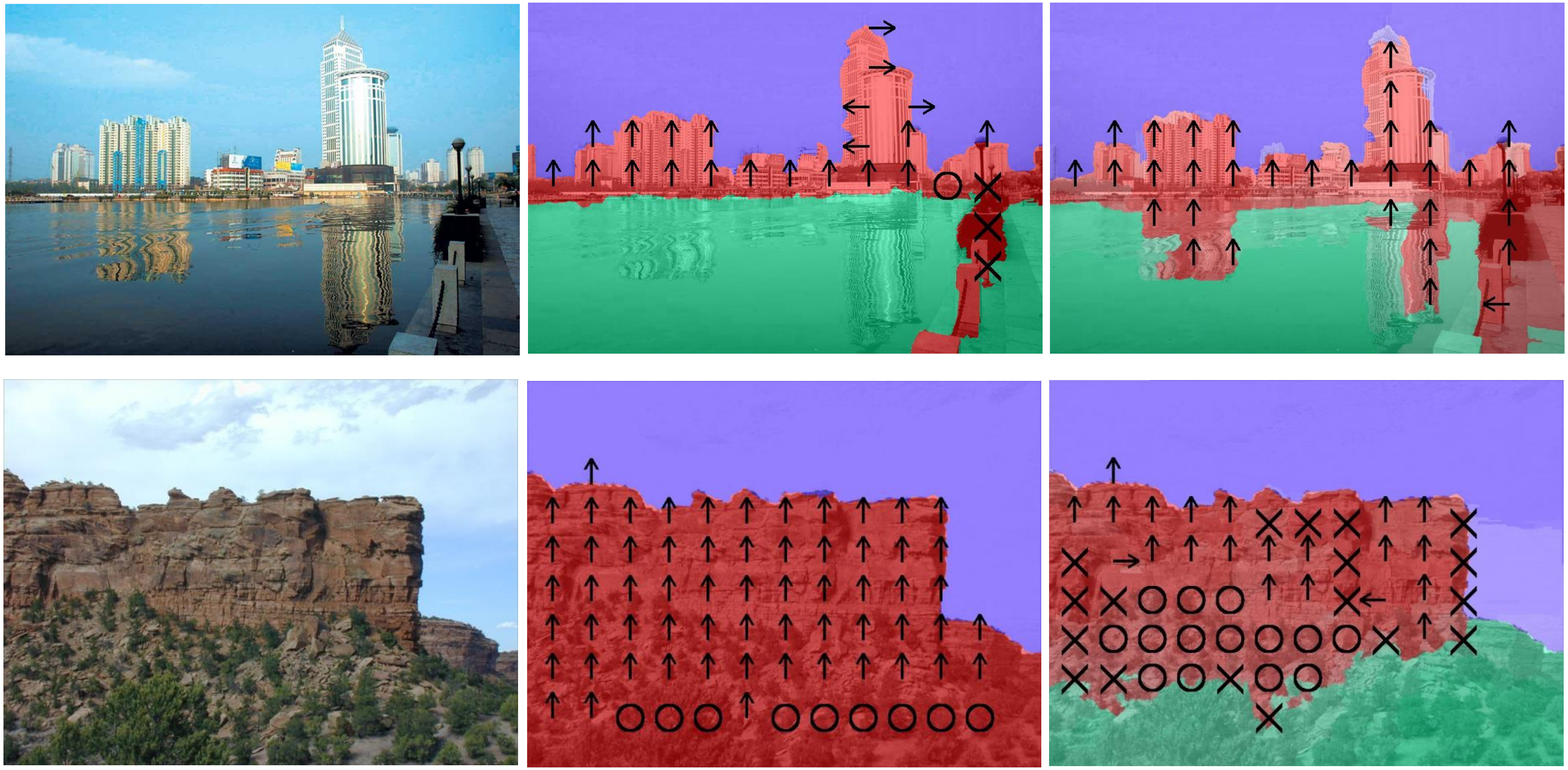


Input Image

Ground Truth

Our Result

Failures: Reflections, Rare Viewpoint



Input Image

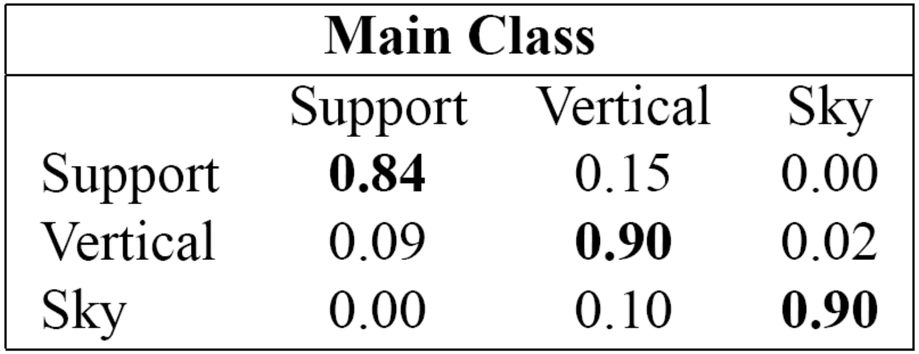
Ground Truth

Our Result

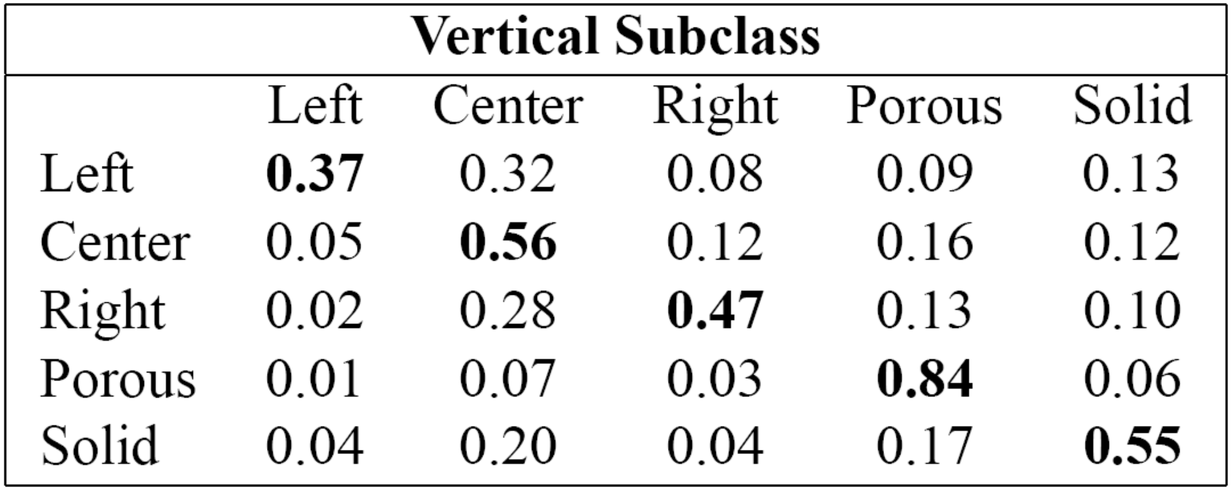
Average Accuracy

Main Class: 88%

Subclasses: 61%



Main Class			
	Support	Vertical	Sky
Support	**0.84**	0.15	0.00
Vertical	0.09	**0.90**	0.02
Sky	0.00	0.10	**0.90**



Vertical Subclass					
	Left	Center	Right	Porous	Solid
Left	**0.37**	0.32	0.08	0.09	0.13
Center	0.05	**0.56**	0.12	0.16	0.12
Right	0.02	0.28	**0.47**	0.13	0.10
Porous	0.01	0.07	0.03	**0.84**	0.06
Solid	0.04	0.20	0.04	0.17	**0.55**

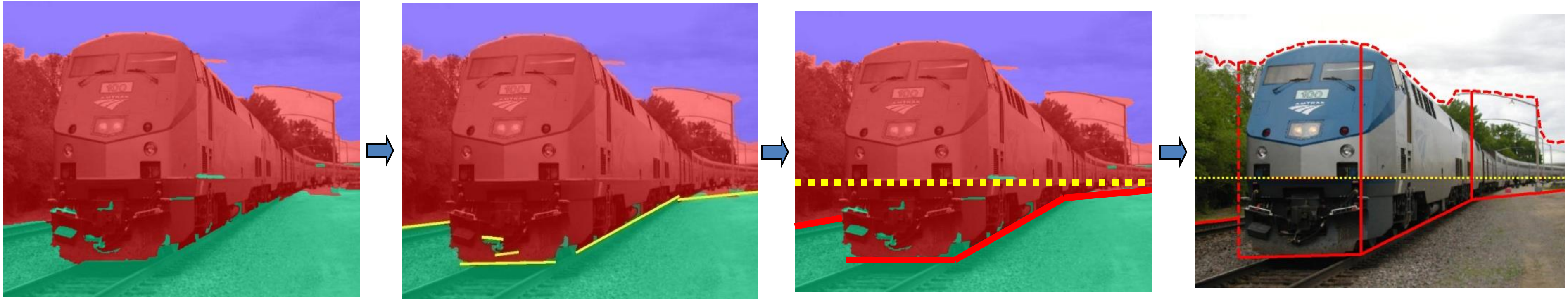
Automatic Photo Popup

Labeled Image

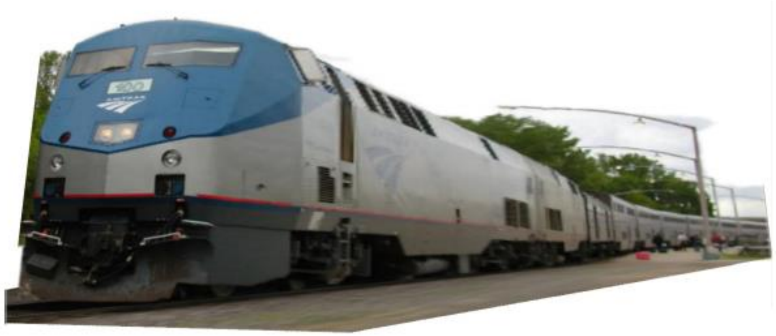
Fit Ground-Vertical Boundary with Line Segments

Form Segments into Polylines

Cut and Fold



Final Pop-up Model

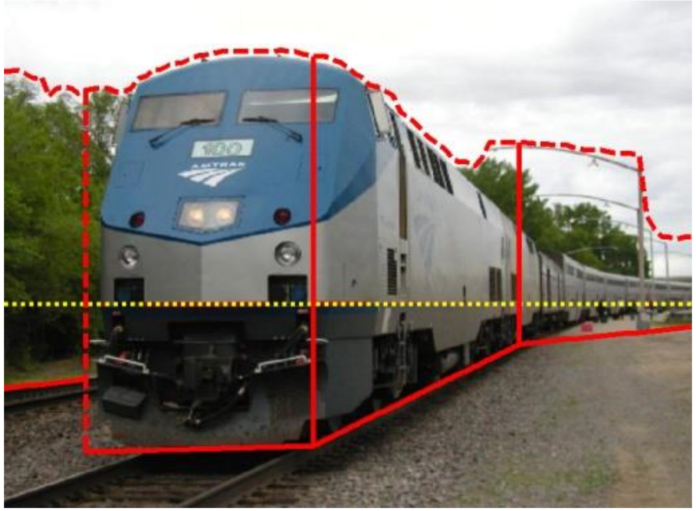


[Hoiem Efros Hebert 2005]

Automatic Photo Pop-up

Slide: Derek Hoiem

Mini-conclusions



- Can learn to predict surface geometry from a single image
- Very rough models, much room for improvement

Things to remember

- Objects should be interpreted in the context of the surrounding scene
  - Many types of context to consider
- Spatial layout is an important part of scene interpretation, but many open problems
  - How to represent space?
  - How to learn and infer spatial models?
- Consider trade-offs of detail vs. accuracy and abstraction vs. quantification