Context

BIL719– Computer Vision

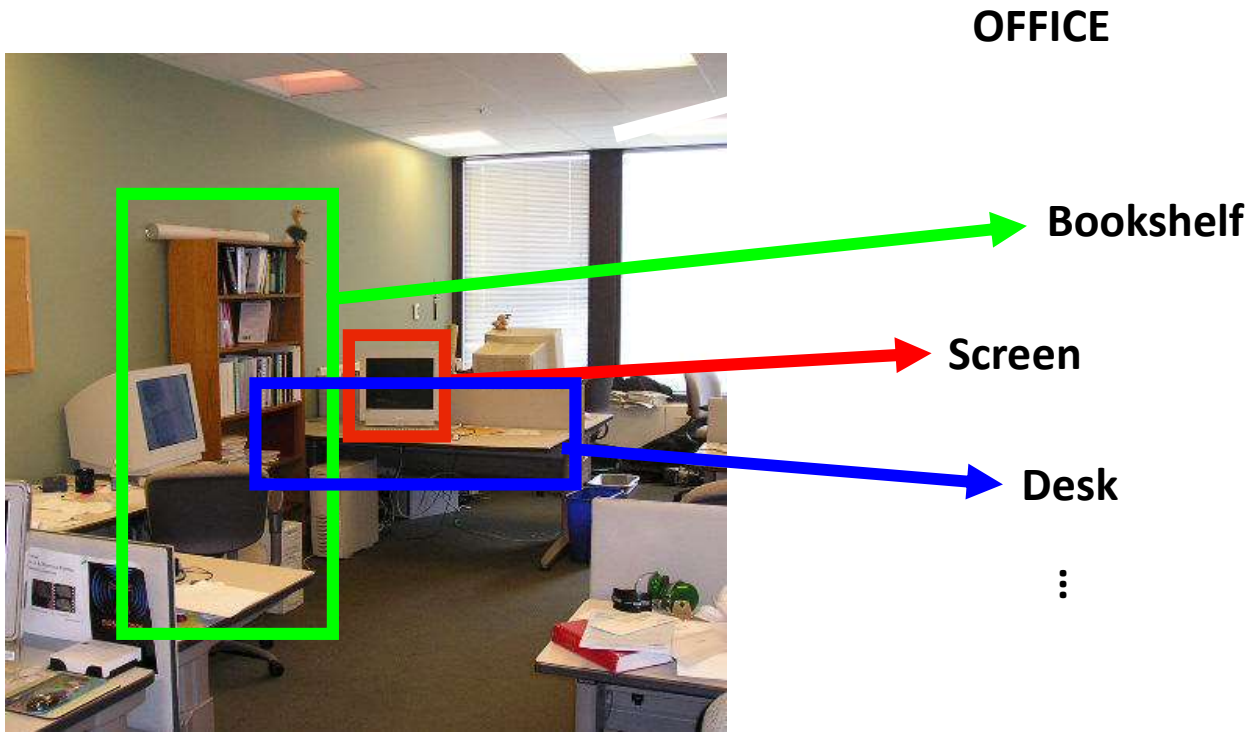
Pinar Duygulu

Hacettepe University

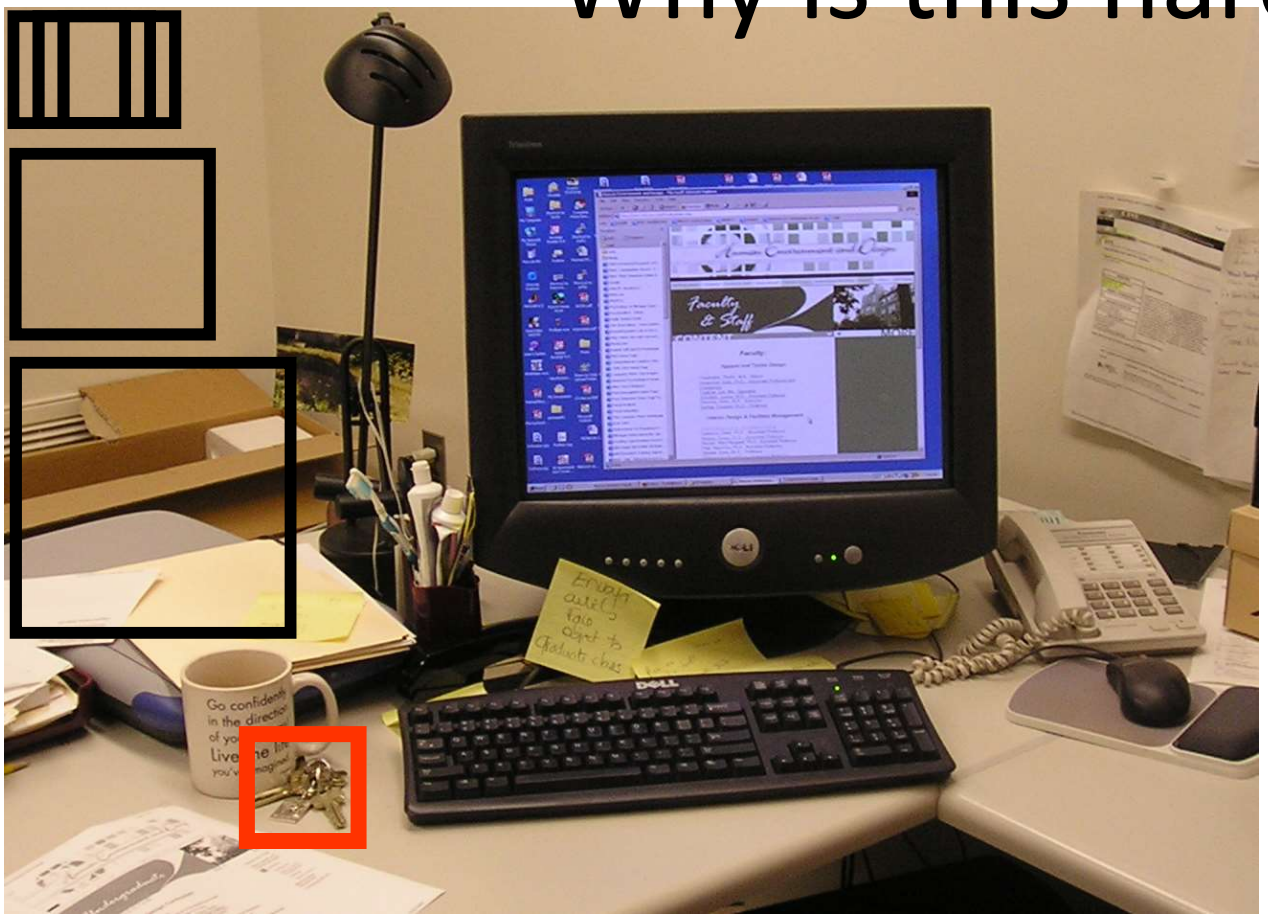
(Source:Antonio Torralba, James Hays)

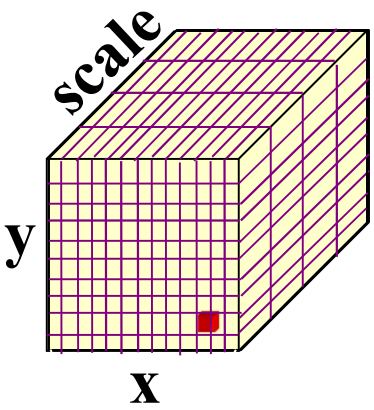
A computer vision goal

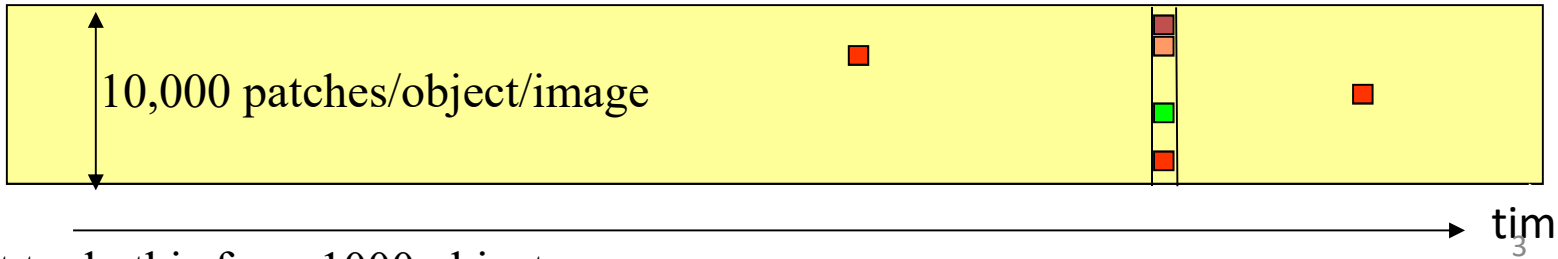
Recognize many different objects under many viewing conditions in unconstrained settings.



Why is this hard?



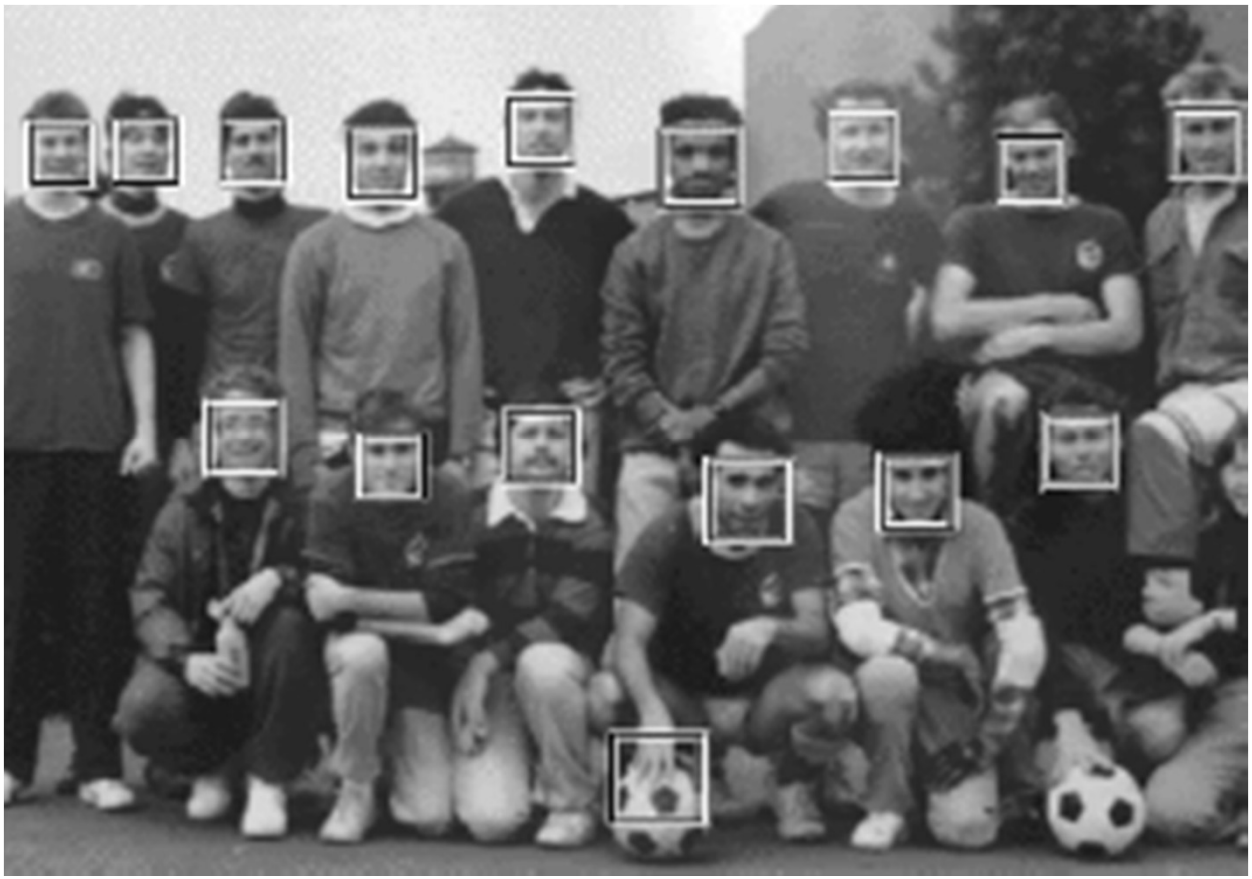




Plus, we want to do this for ~ 1000 objects

1,000,000 images/day

The face detection age



- The representation and matching of pictorial structures Fischler, Elschlager (1973).
- Face recognition using eigenfaces M. Turk and A. Pentland (1991).
- Human Face Detection in Visual Scenes - Rowley, Baluja, Kanade (1995)
- Graded Learning for Object Detection - Fleuret, Geman (1999)
- Robust Real-time Object Detection - Viola, Jones (2001)
- Feature Reduction and Hierarchy of Classifiers for Fast Object Detection in Video Images - Heisele, Serre, Mukherjee, Poggio (2001)
- ….

“Head in the coffee beans problem”

Can you find the head in this image?

“Head in the coffee beans problem”

Can you find the head in this image?

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.

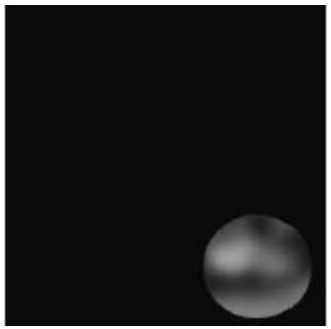
Context provides clues for function

- What is this?

- Now can you tell?

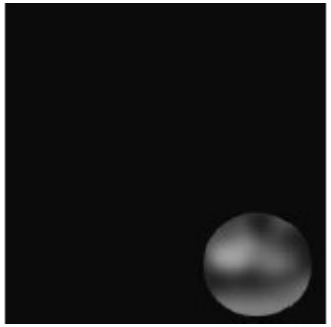
Sometimes context is *the* major component of recognition

- What is this?



Sometimes context is *the* major component of recognition

- What is this?



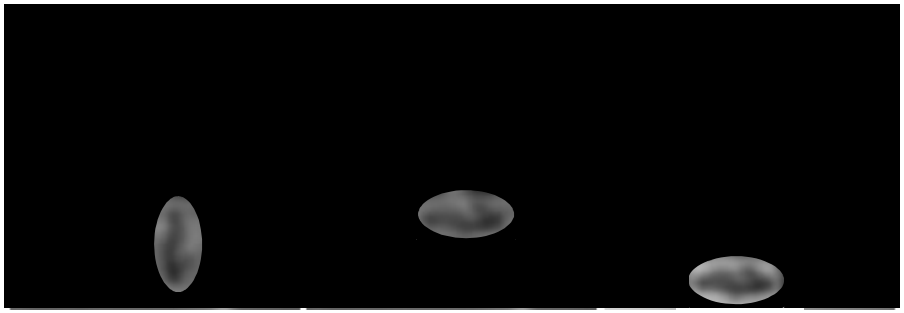
- Now can you tell?

More Low-Res

- What are these blobs?

More Low-Res

- The same pixels! (a car)



Some symptoms of standard approaches

Just objects is not enough

The detector challenge: by looking at the output of a detector on a random set of images, can you guess which object is it trying to detect?

What object is detector trying to detect?

The detector challenge: by looking at the output of a detector on a random set of images, can you guess which object is it trying to detect?

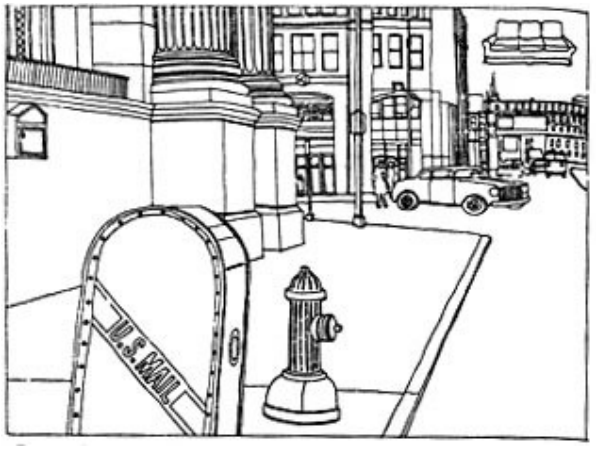
1. chair, 2. table, 3. road, 4. road, 5. table, 6. car, 7. keyboard.

What are the hidden objects?

What are the hidden objects?

Biederman 1982

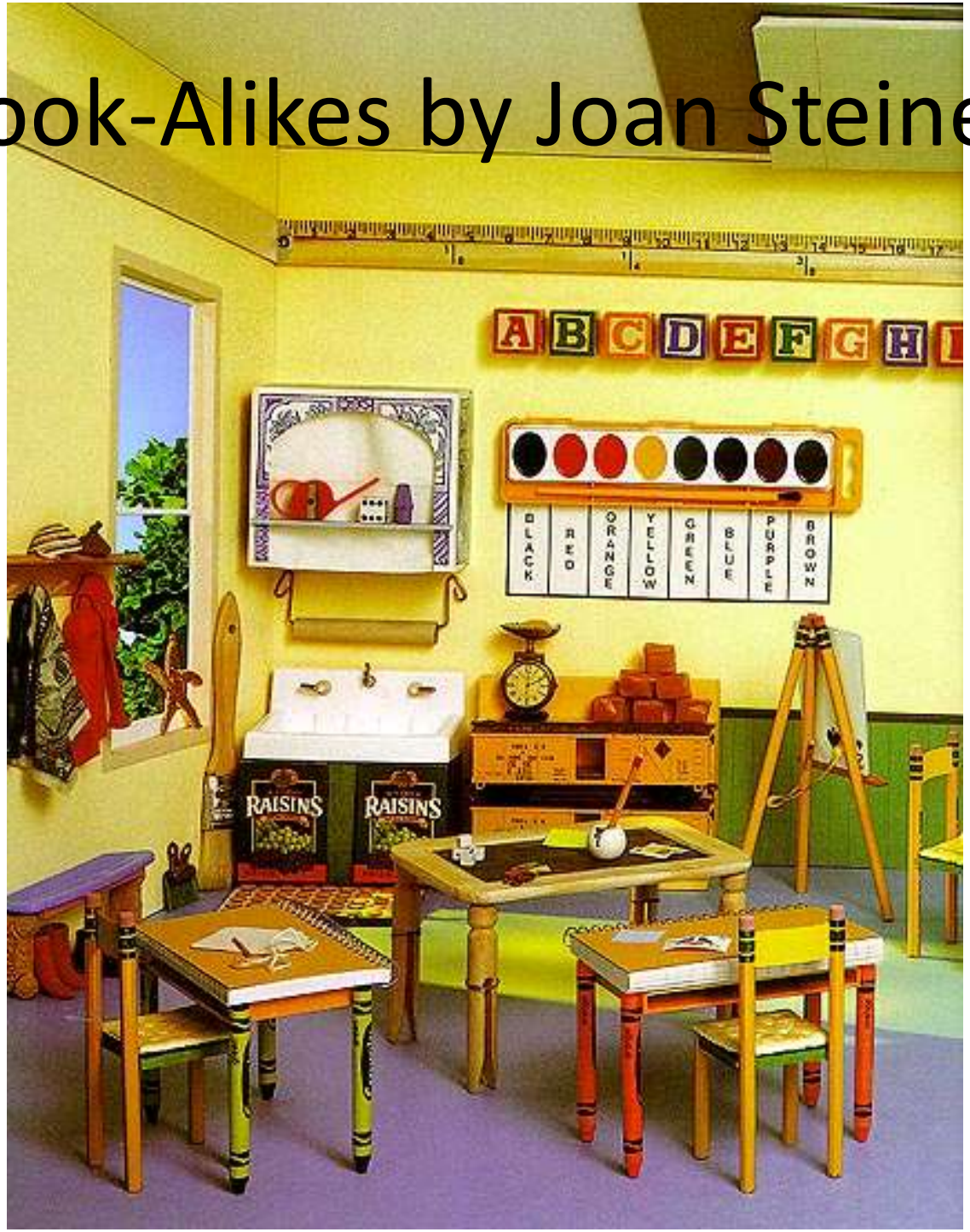
- Pictures shown for 150 ms.
- Objects in appropriate context were detected more accurately than objects in an inappropriate context.
- Scene consistency affects object detection.



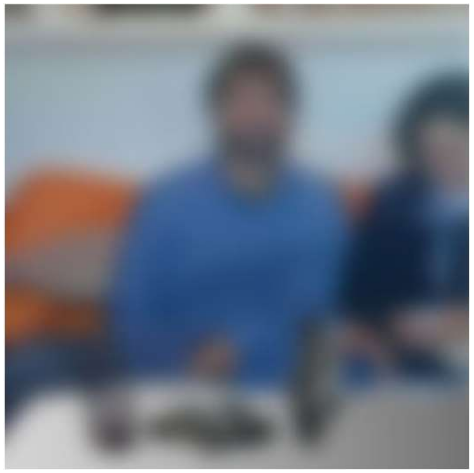
Look-Alikes by Joan Steiner

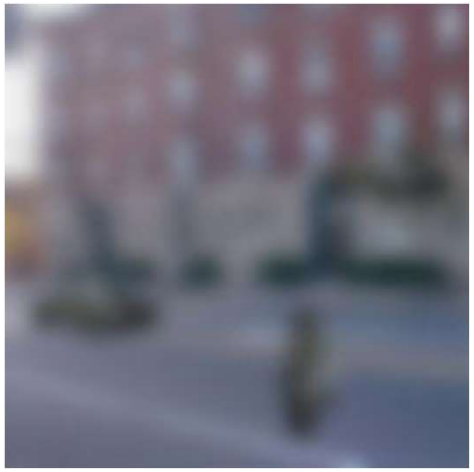
Even in high resolution, we can not shut down contextual processing and it is hard to recognize the true identities of the elements that compose this scene.

Look-Alikes by Joan Steiner

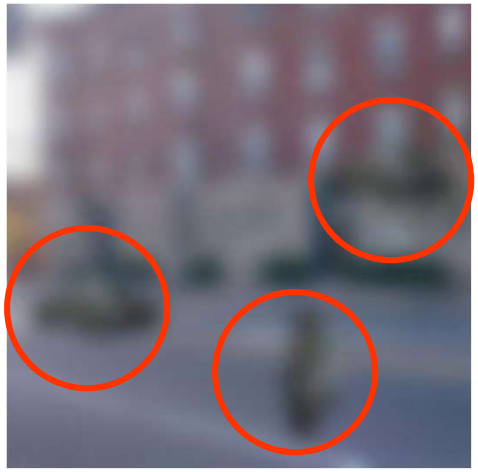


The multiple personalities of a blob

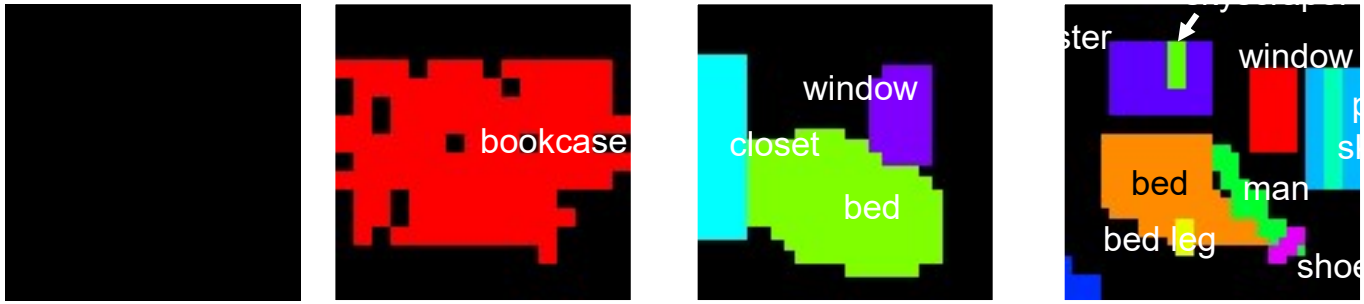





The multiple personalities of a blob



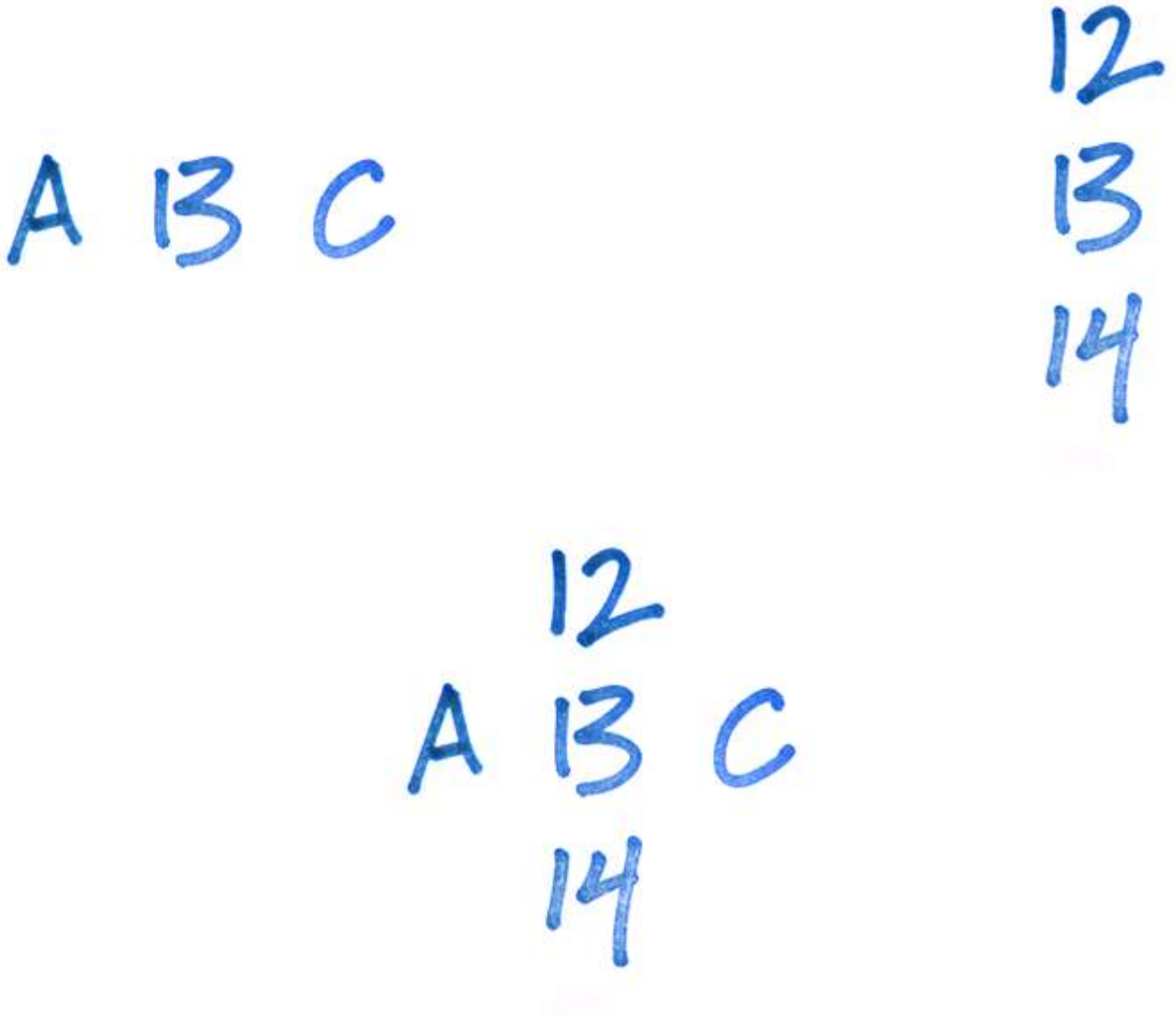
Recognition with low resolution



Disambiguation

Disambiguation

Disambiguation



Why is context important?

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

- Context defines what an unexpected event is

Global precedence

Forest Before Trees: The Precedence of Global Features in Visual Perception
Navon (1977)

$$p(O \mid I) \alpha p(I|O)\ p(O)$$

Object model

Context model

Full joint

Scene model

Approx. joint

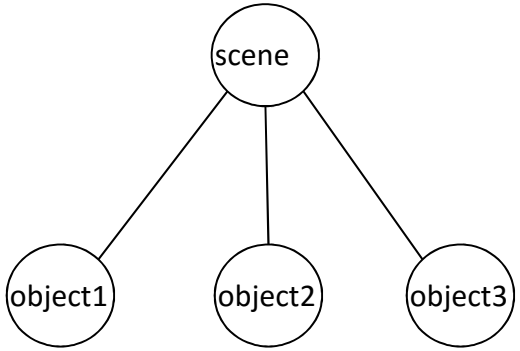
$$p(O) = \sum_{s} \prod_{i} p(Oi|S=s)\ p(S=s)$$

office

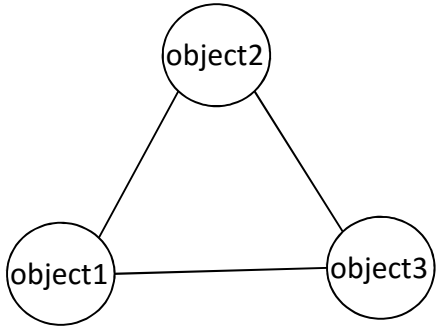
street

Context models

Independent model



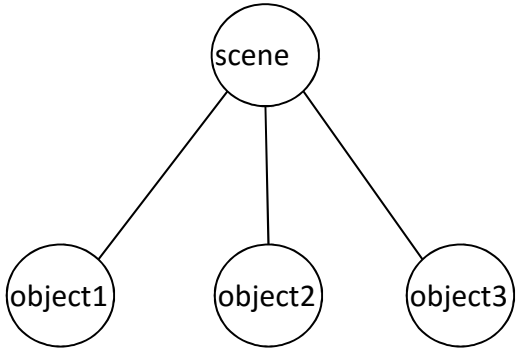
Objects are correlated via the scene



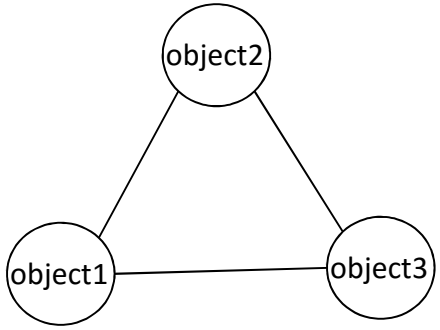
Dependencies among objects

Context models

Independent model

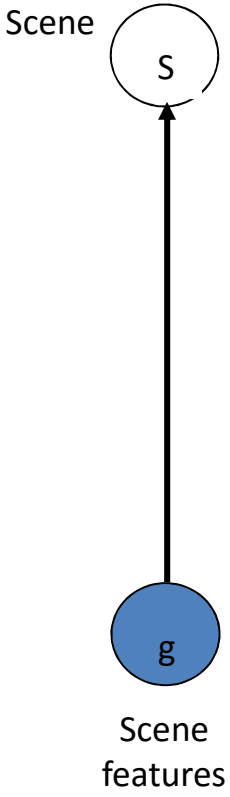


Objects are correlated via the scene



Dependencies among objects

Scene recognition without object recognition



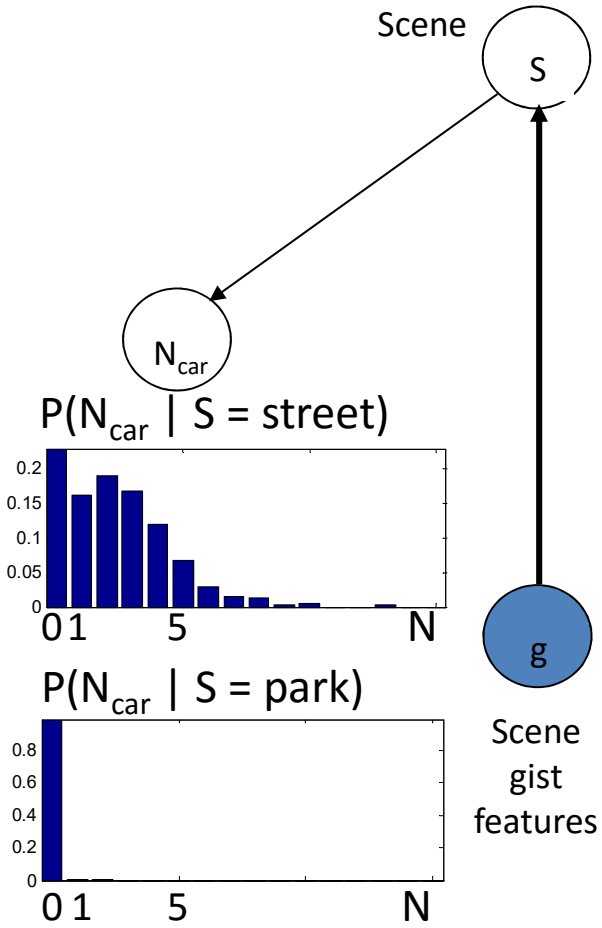
Murphy, Torralba, Freeman; NIPS 2003. Torralba, Murphy, Freeman, CACM 2010.

Application of object detection for image retrieval






An integrated model of Scenes, Objects, and Parts

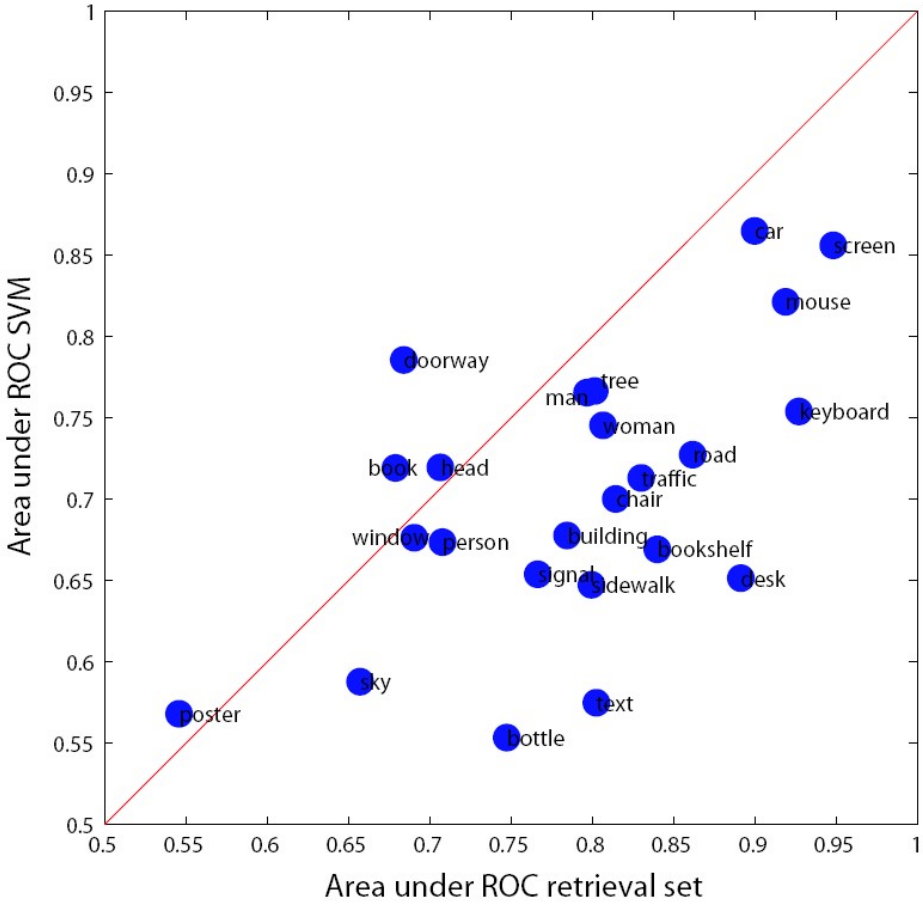


Murphy, Torralba, Freeman; NIPS 2003. Torralba, Murphy, Freeman, CACM 2010.

Object retrieval: scene features vs. detector

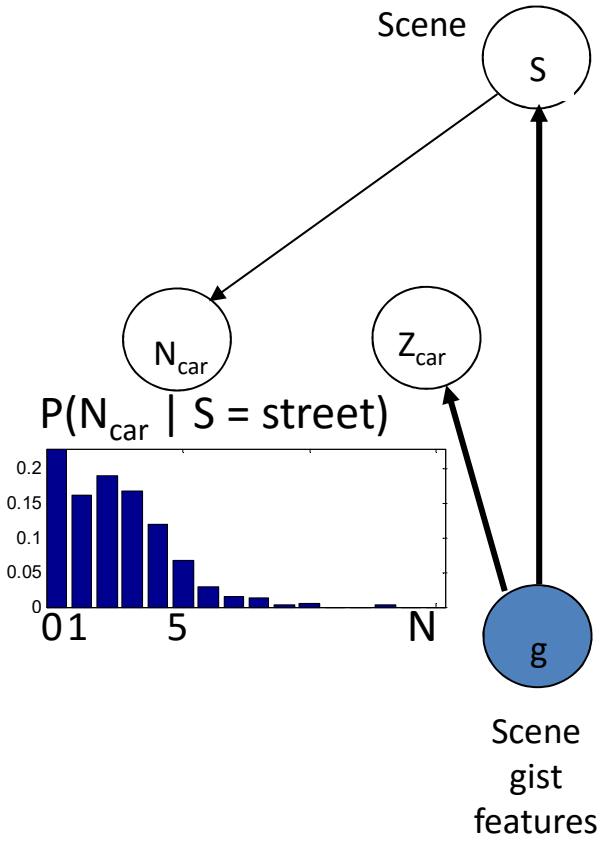
Results using the keyboard detector alone

Results using both the detector and the global scene features



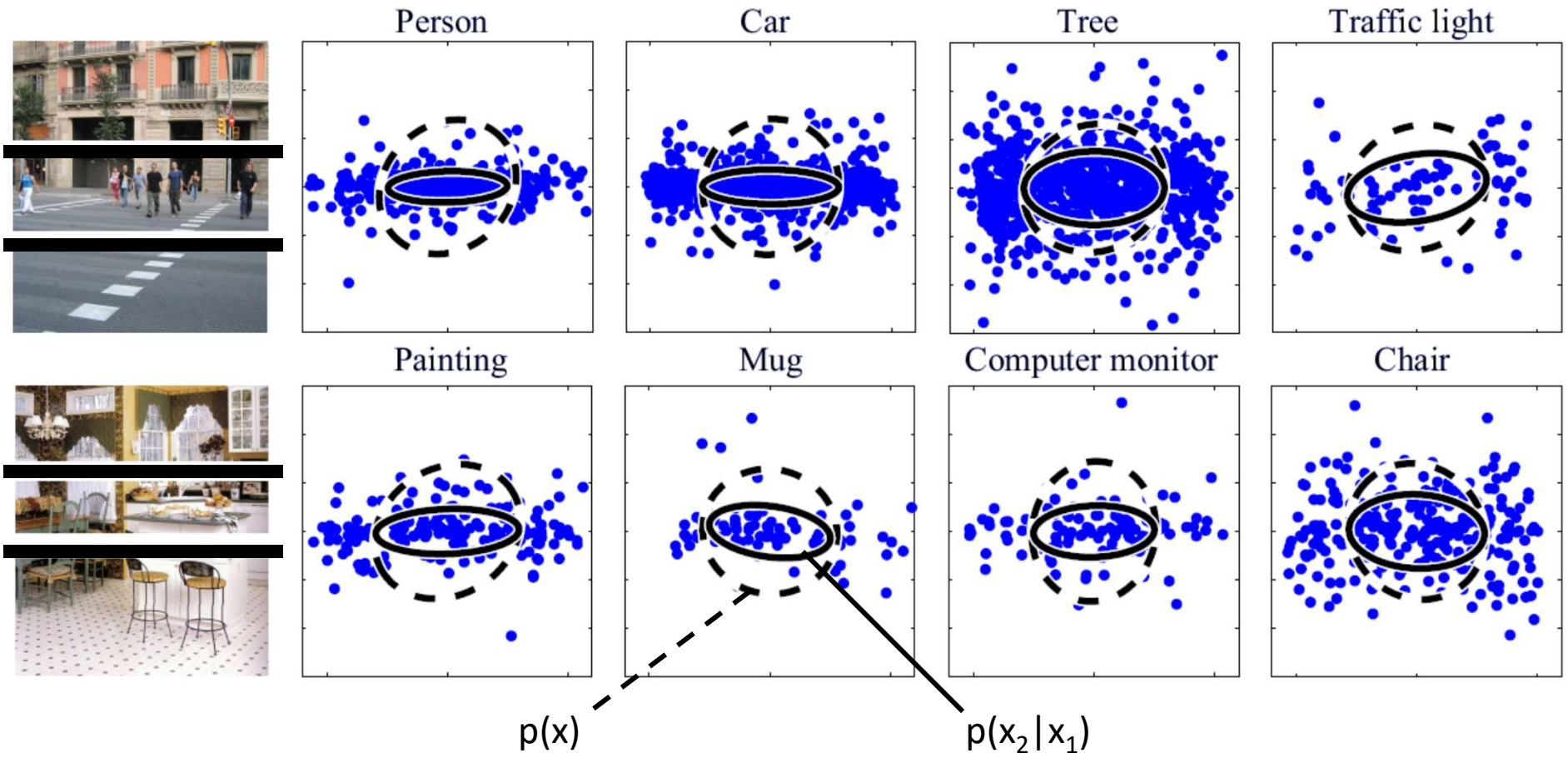
Murphy, Torralba, Freeman; NIPS 2003. Torralba, Murphy, Freeman, CACM 2010.

Context driven object detection



The layered structure of scenes

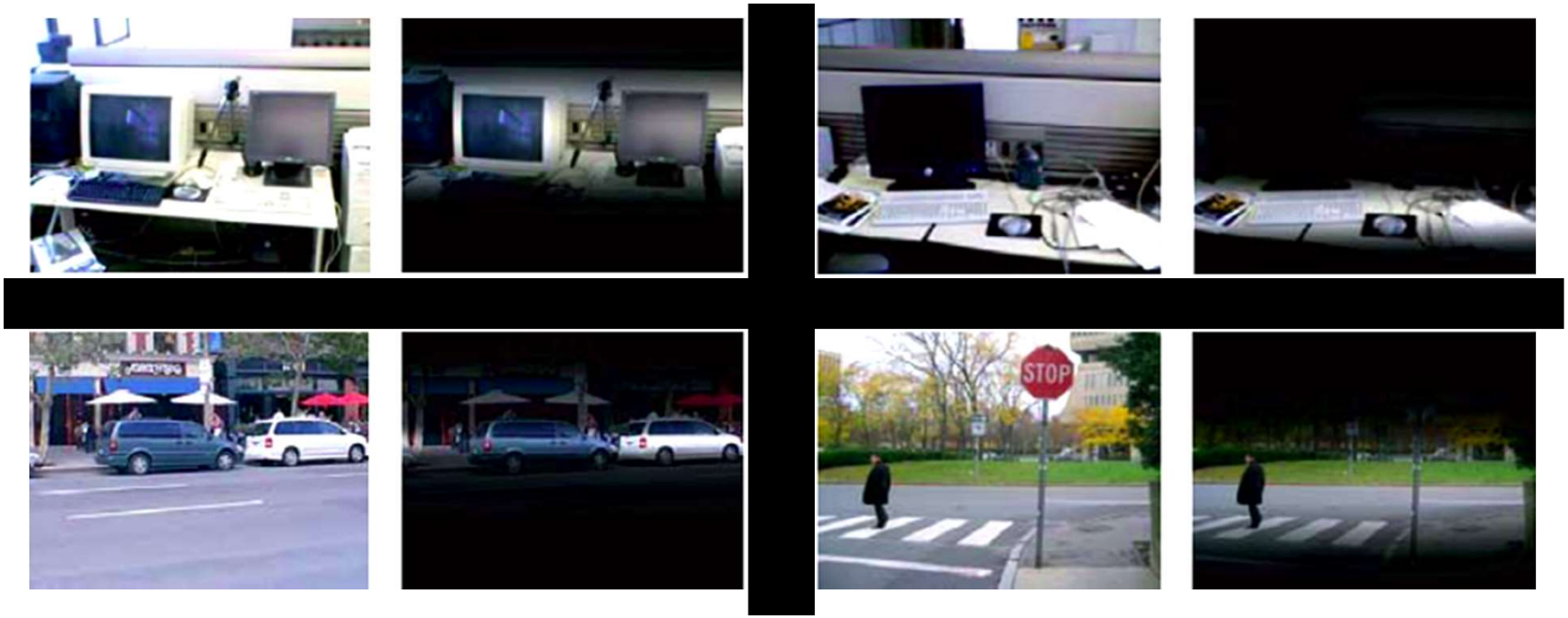
Assuming a human observer standing on the ground



In a display with multiple targets present, the location of one target constraints the 'y' coordinate of the remaining targets, but not the 'x' coordinate.

Torralba, Oliva, Castelhano, Henderson. 2006

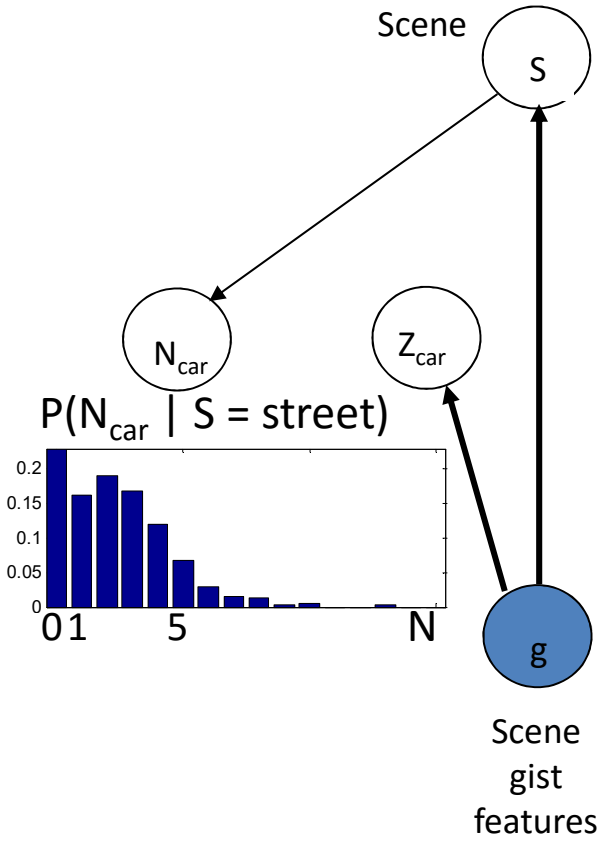
Car detection without a car detector


STOP

Detecting faces without a face detector

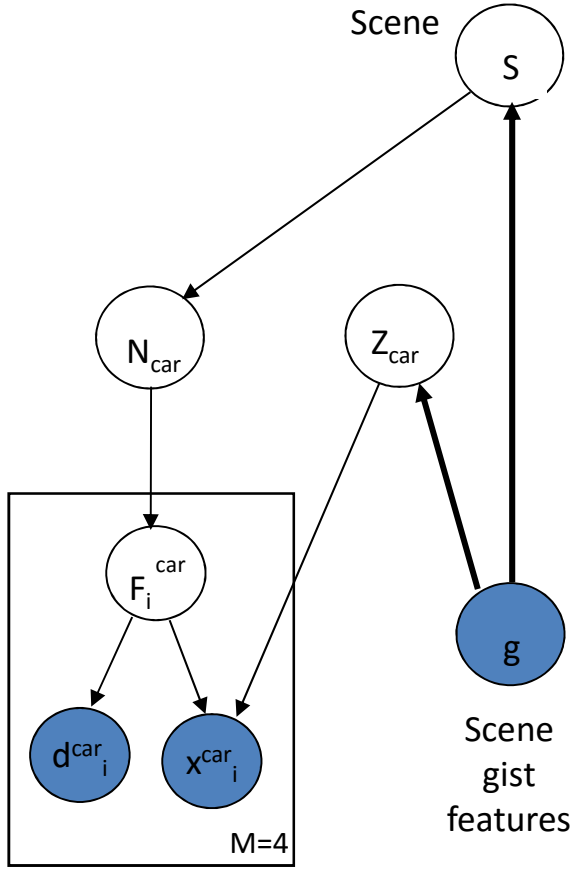
Torralba & Sinha, 01; Torralba, 03

Context driven object detection

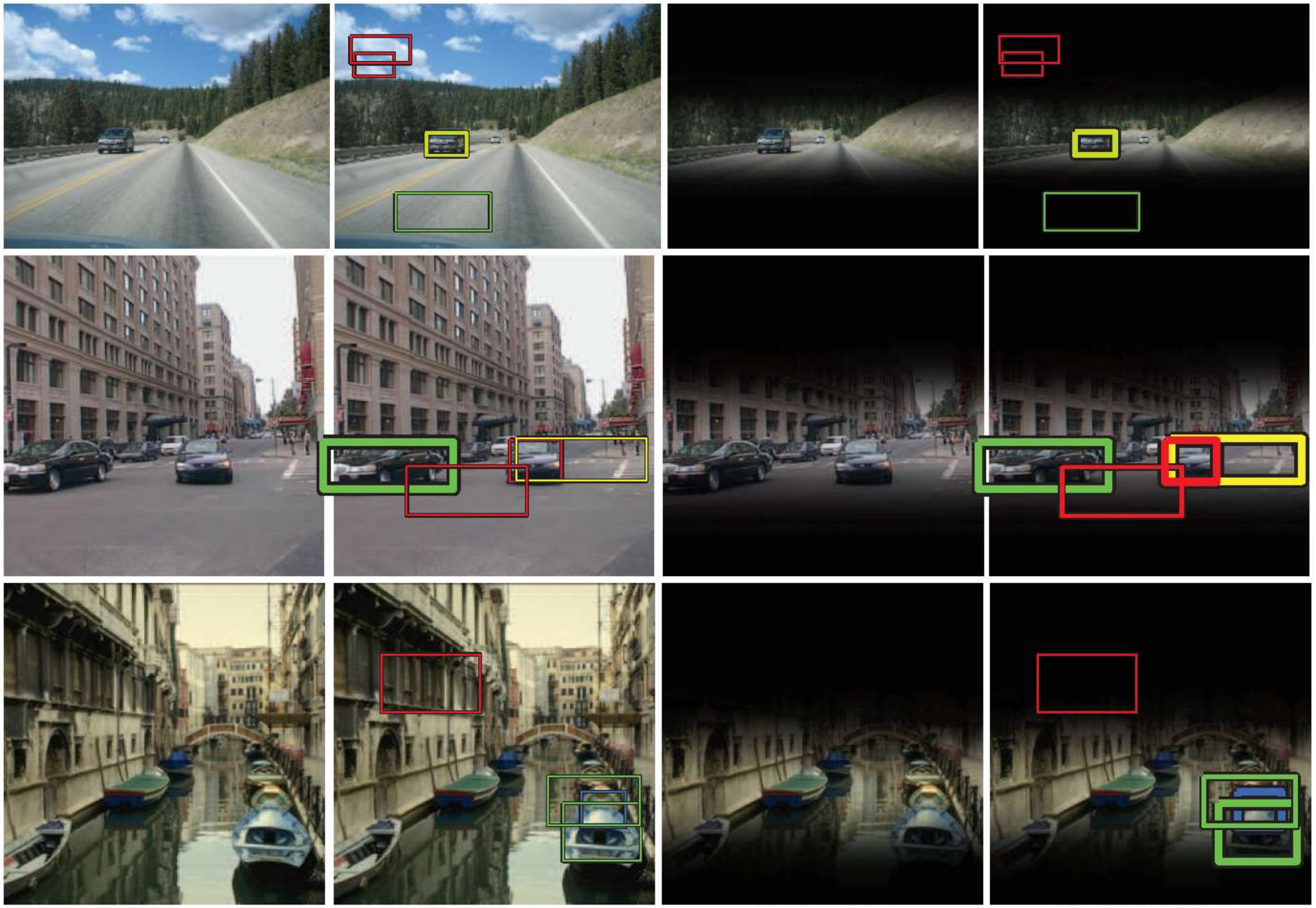


Murphy, Torralba, Freeman; NIPS 2003. Torralba, Murphy, Freeman, CACM 2010.

An integrated model of Scenes, Objects, and Parts



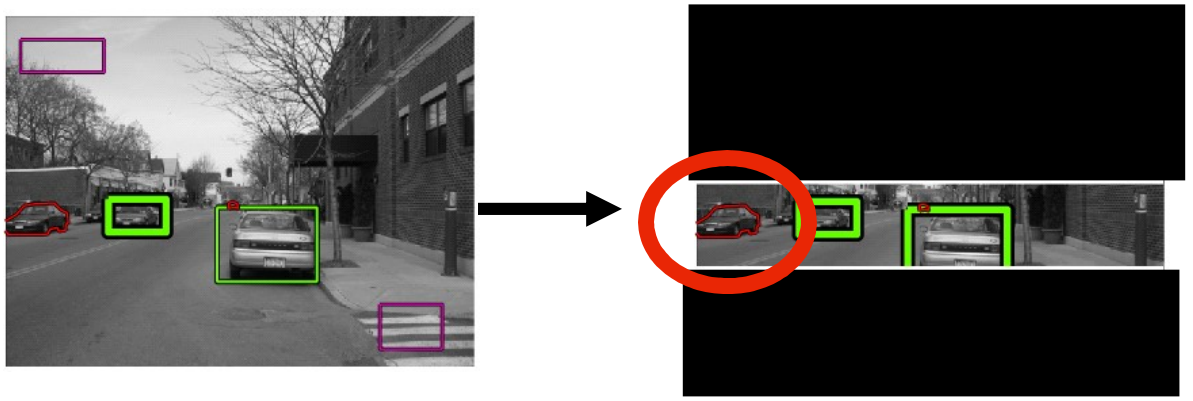
Murphy, Torralba, Freeman; NIPS 2003. Torralba, Murphy, Freeman, CACM 2010.

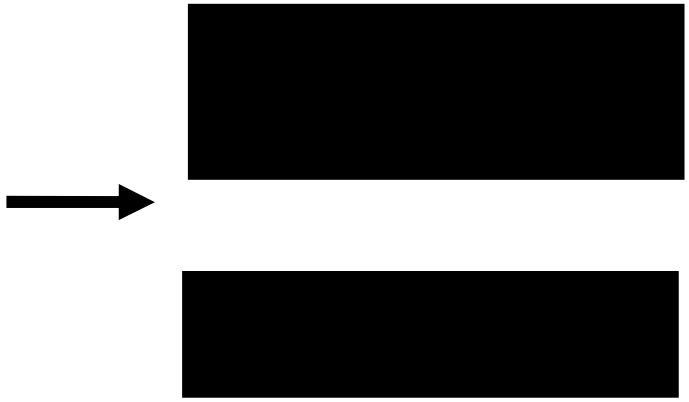


a) input image b) car detector output c) location priming c) integrated model output

Failures

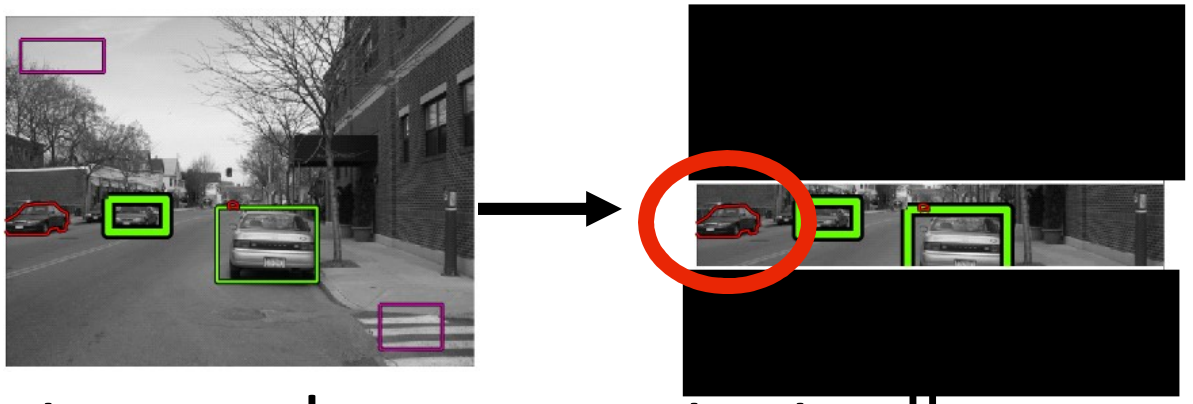
- If the detector fails… context can not help



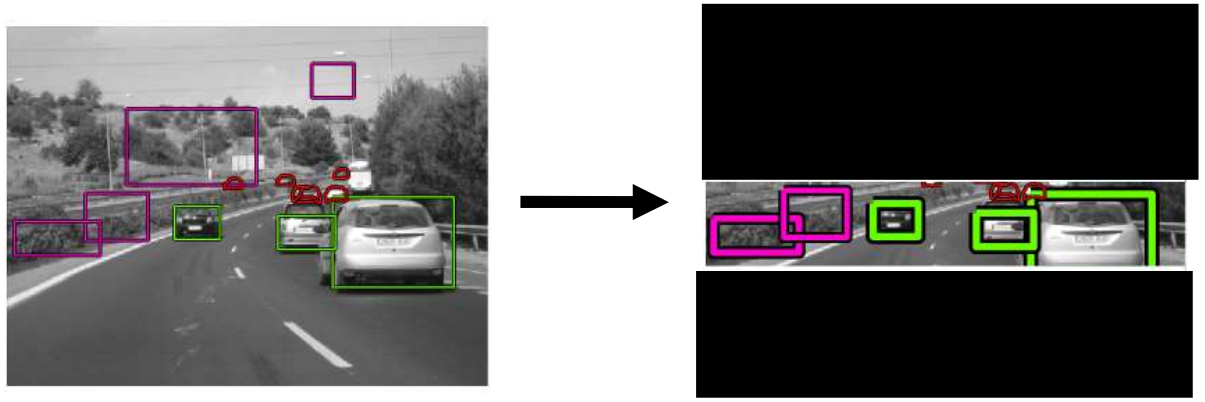


Failures

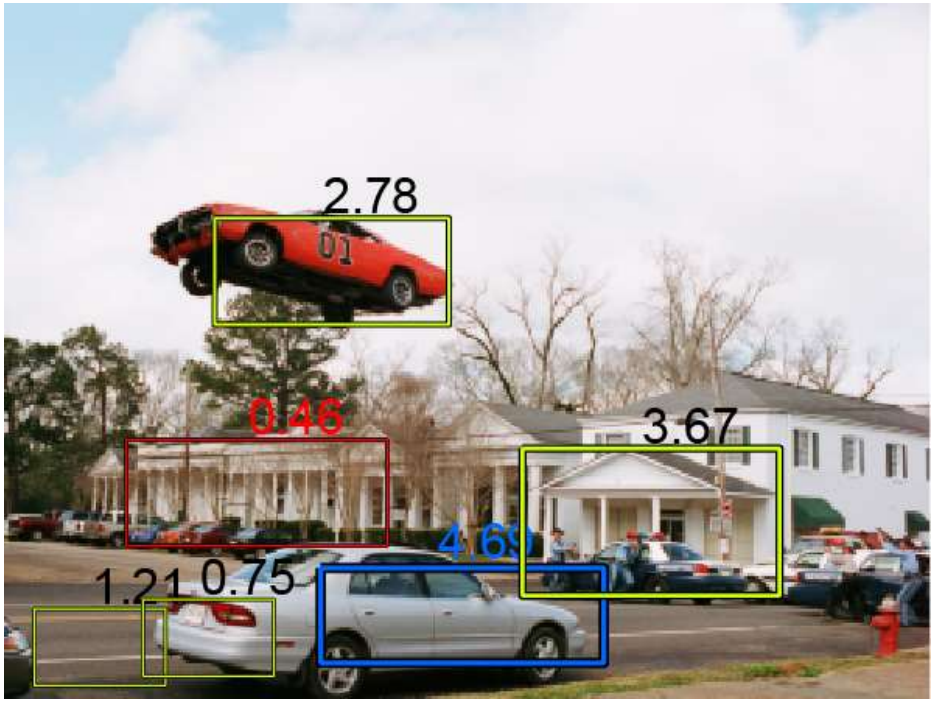
- If the detector fails… context can not help

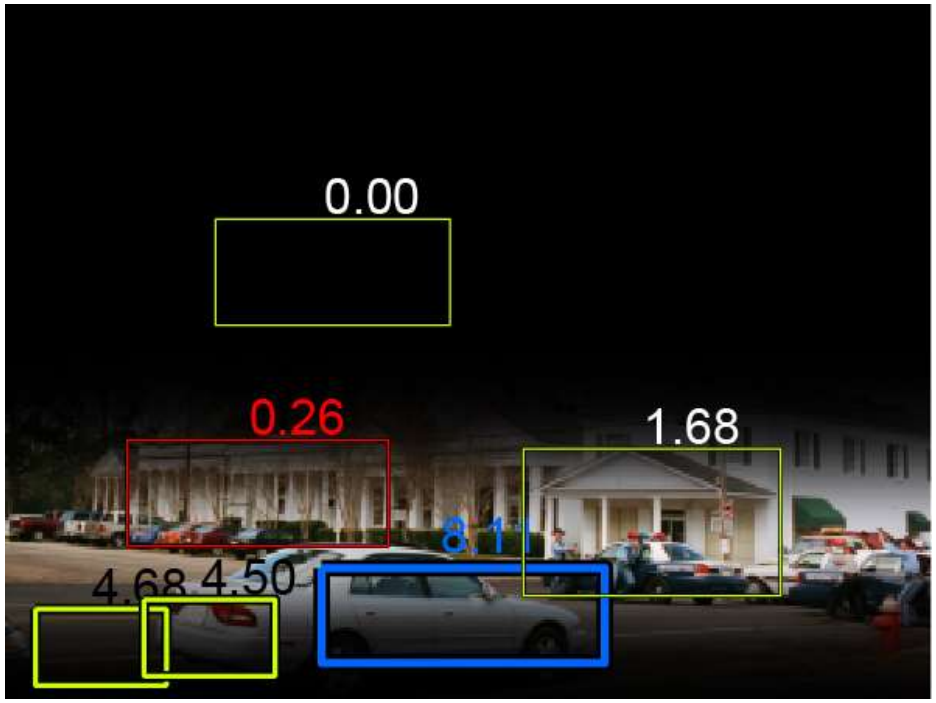


- If the detector produces a contextually coherent false alarm, context will increase the error.



A car out of context …





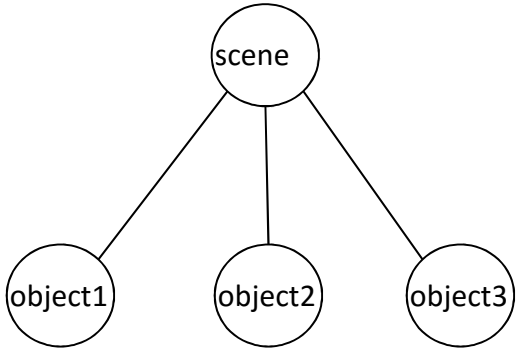
Who needs context anyway?

We can recognize objects even out of context

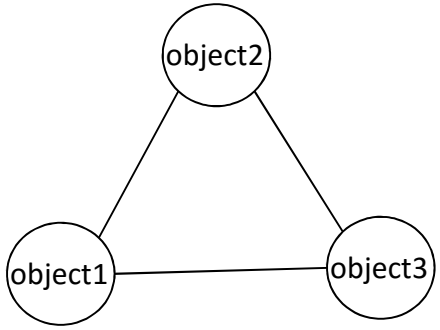
Banksy

Context models

Independent model



Objects are correlated via the scene



Dependencies among objects

1) Generate candidate objects (run a detector, or segmentation)

M possible object labels
N regions

Label: c_k = [1…M] with k = [1…N]
Scores: s_k = vector length M

2) For each candidate, get a list of possible interpretations with their probabilities

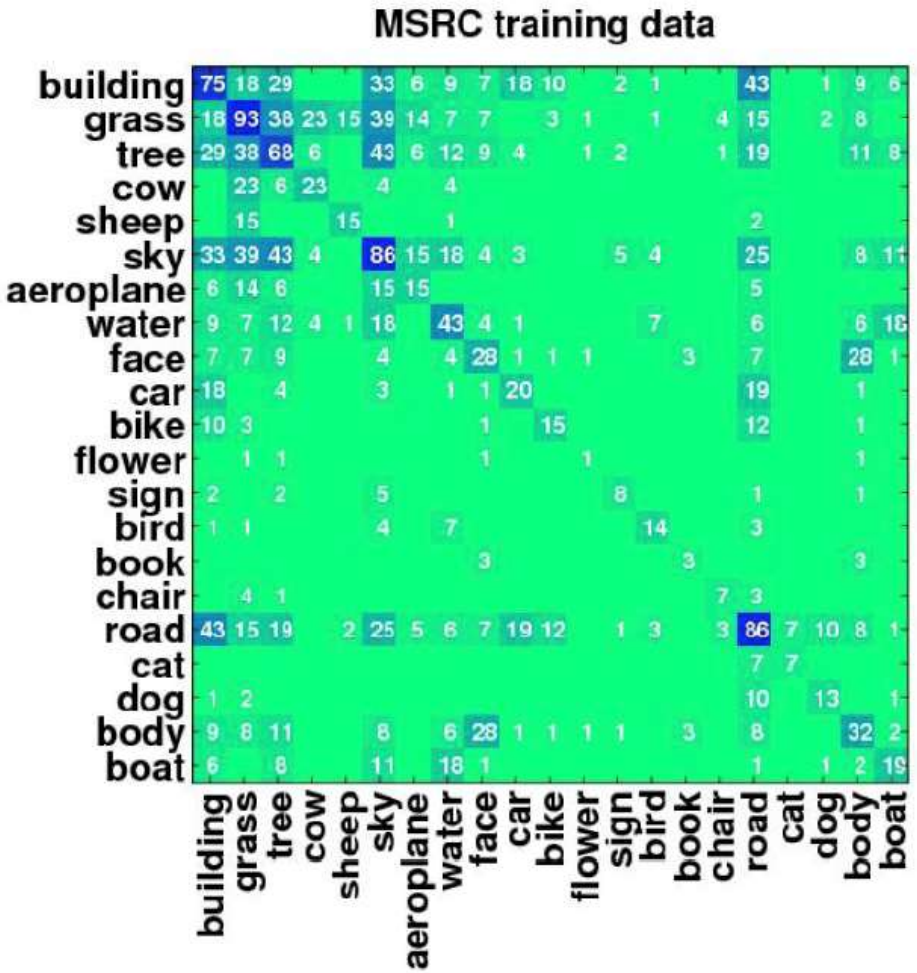
$$p(c_k = m \mid s_k)$$

3) Goal: to assign labels c_k to each candidate so that they are in contextual agreement. We want to optimize the joint probability of all the labels:

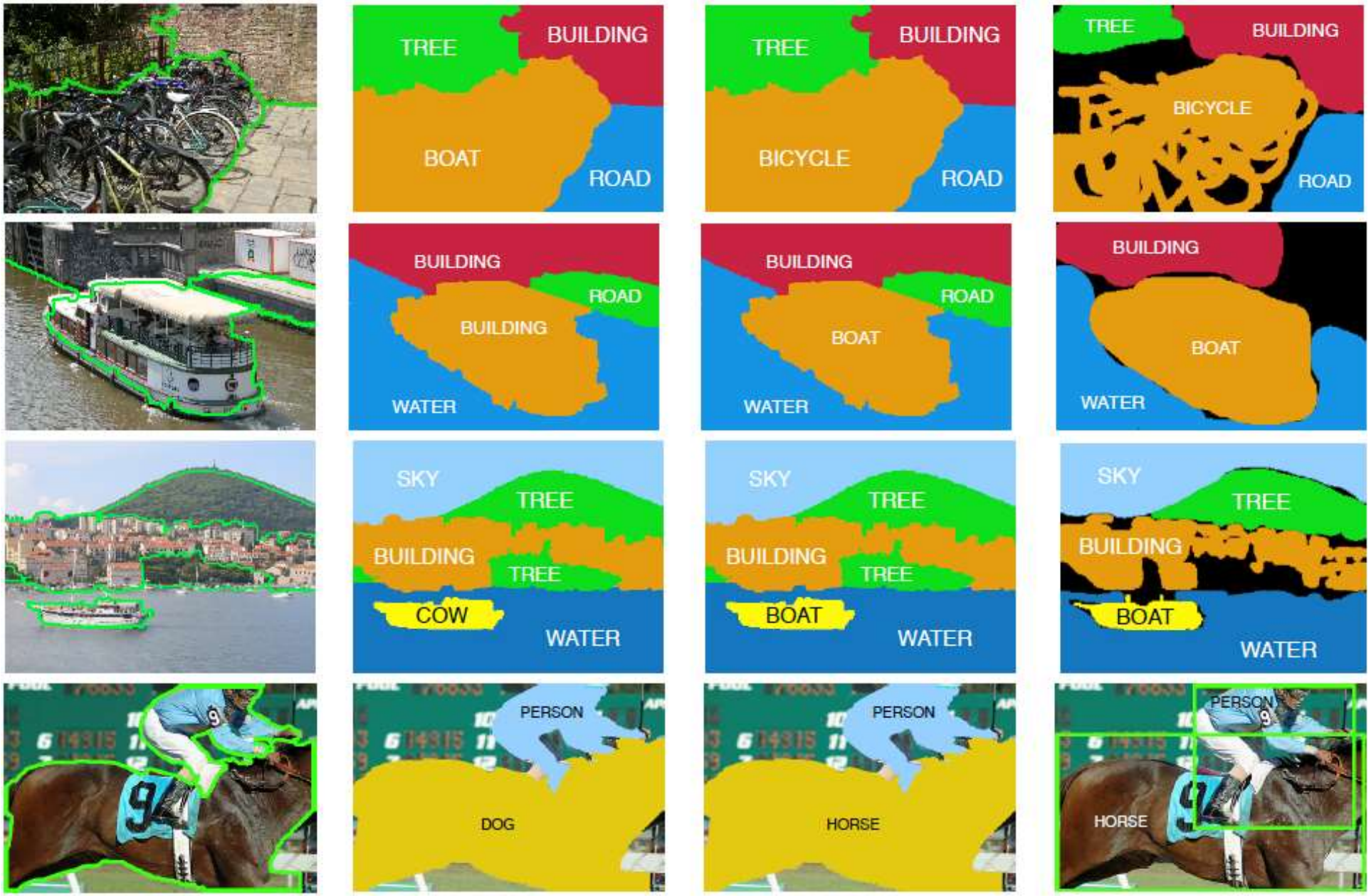
$$p(c_1 = m_1, \ldots, c_N = m_N \mid s_1, \ldots, s_N)$$

A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora and S. Belongie. Objects in Context. ICCV 2007

$\Phi(c_i=m_i, c_j=m_j)$ = co-ocurrence matrix on training set (count how many times two objects appear together).



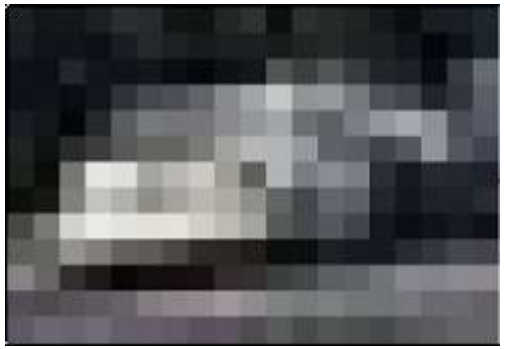
A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora and S. Belongie. Objects in Context. ICCV 2007

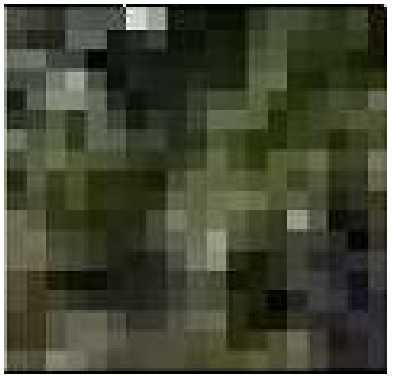


A. Rabinovich, A. Vedaldi, C. Galleguillos, E. Wiewiora and S. Belongie. Objects in Context. ICCV 2007

Spatial layout is especially important

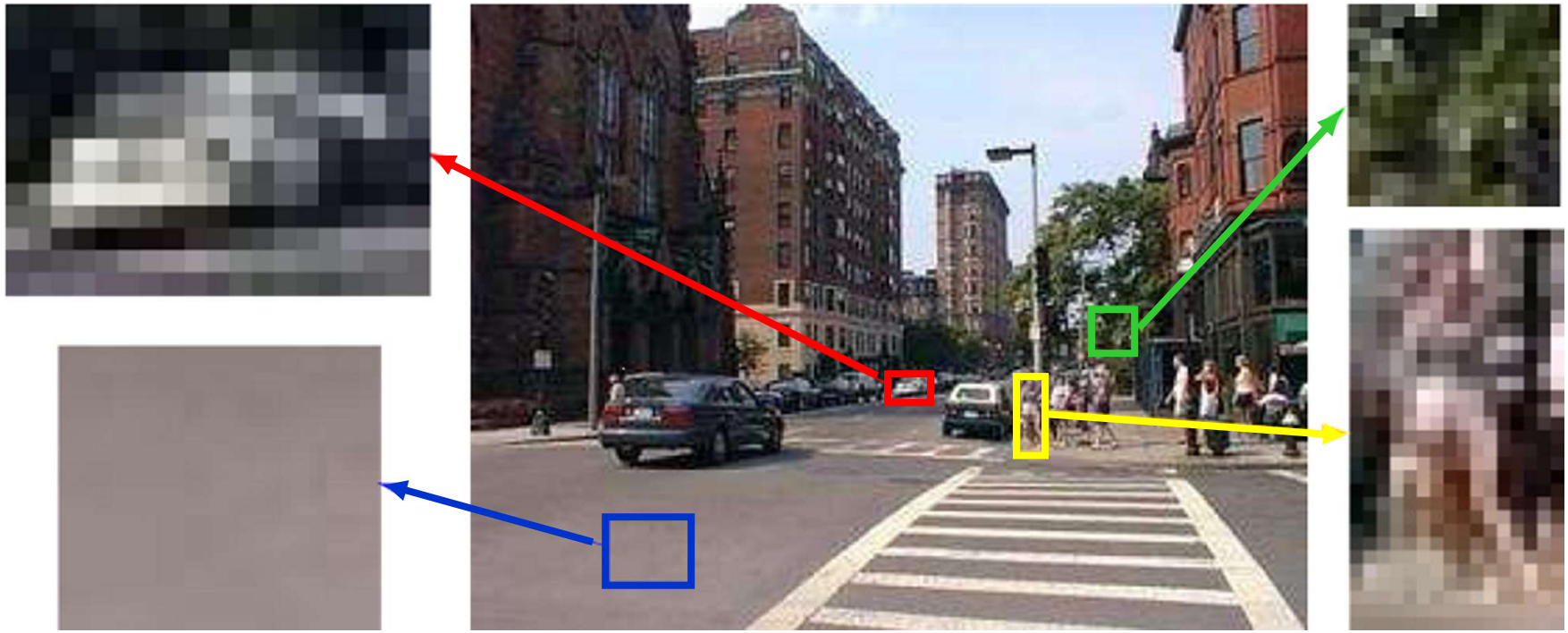
1. Context for recognition





Spatial layout is especially important

1. Context for recognition



Slide: Derek Hoiem

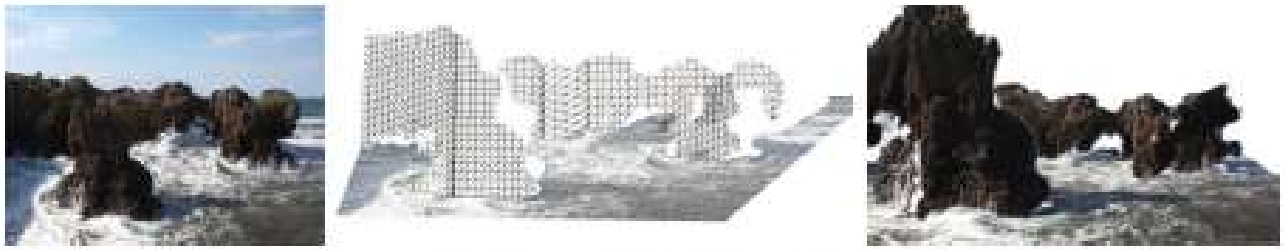
Spatial layout is especially important

1. Context for recognition
2. Scene und

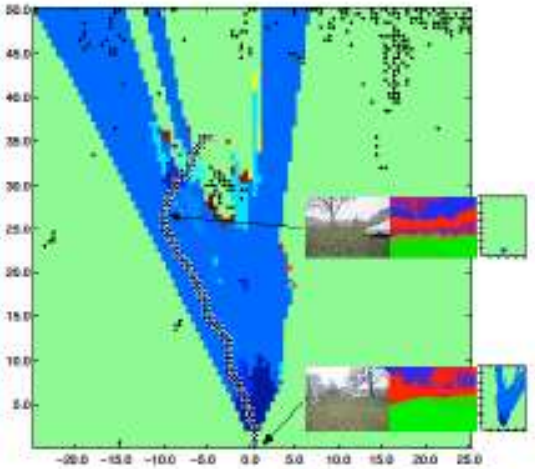
Slide: Derek Hoiem

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

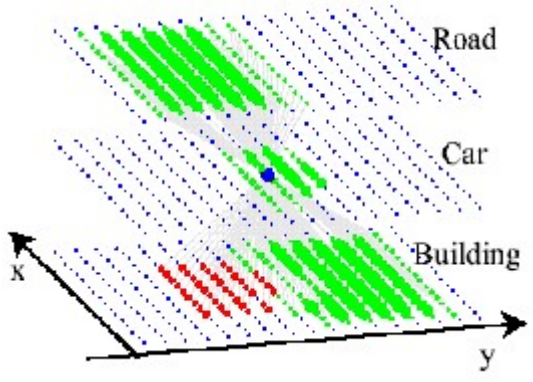


Robot Navigation: Path Planning

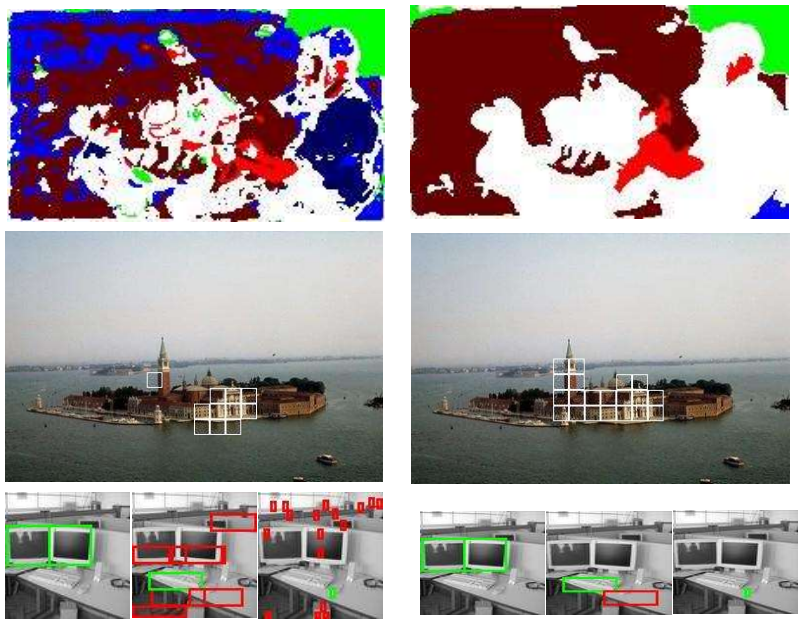
Spatial Layout: 2D vs. 3D

Slide: Derek Hoiem

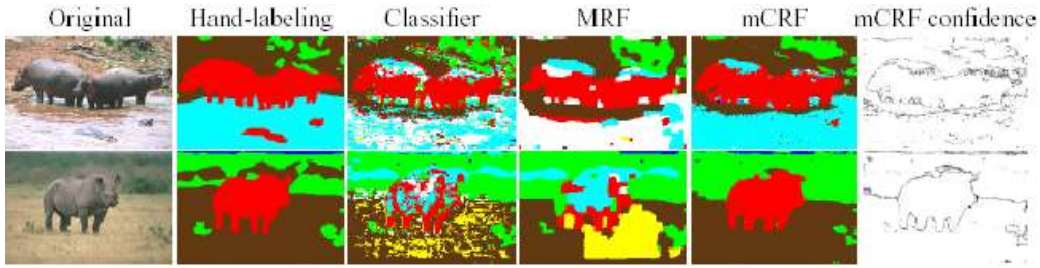
Context in Image Space



[Torralba Murphy Freeman 2004]



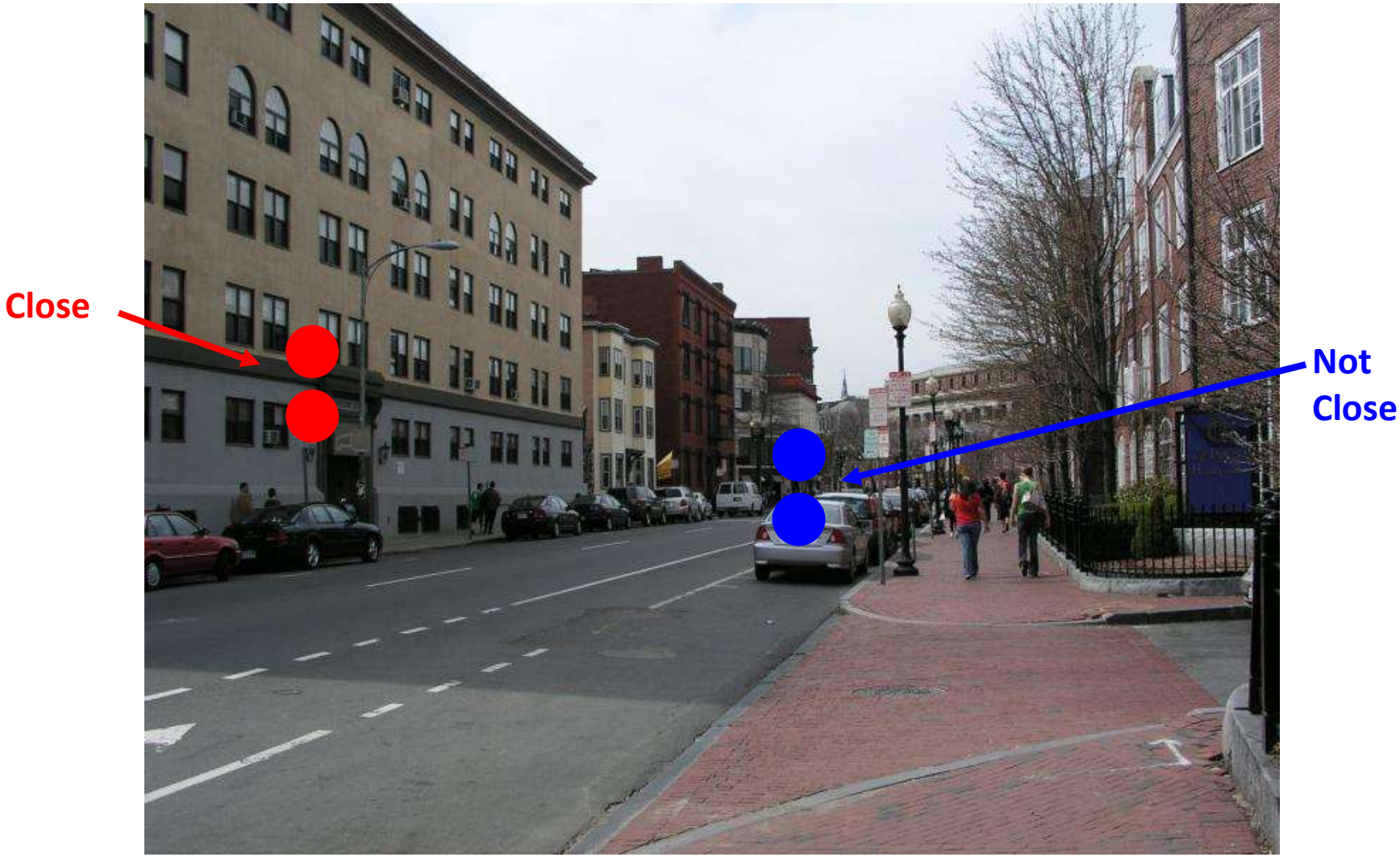
[Kumar Hebert 2005]



[He Zemel Cerreira-Perpiñán 2004]

Slide: Derek Hoiem

But object relations are in 3D…



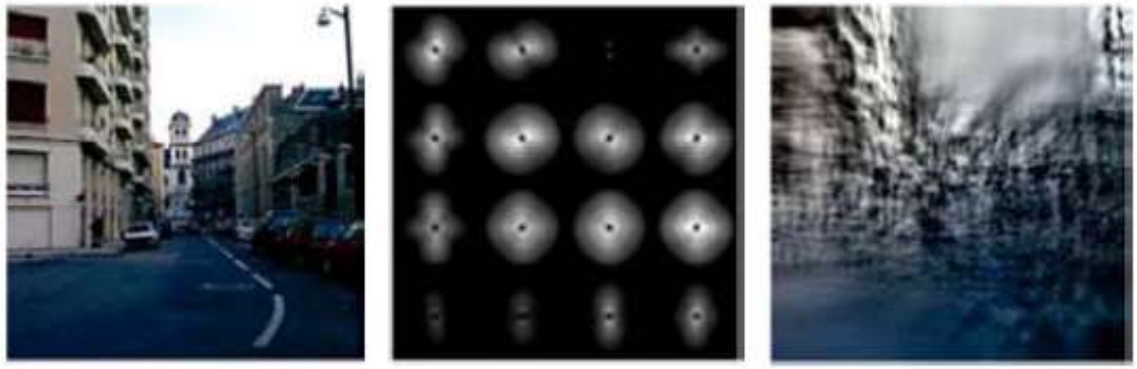
Slide: Derek Hoiem

How to represent scene space?

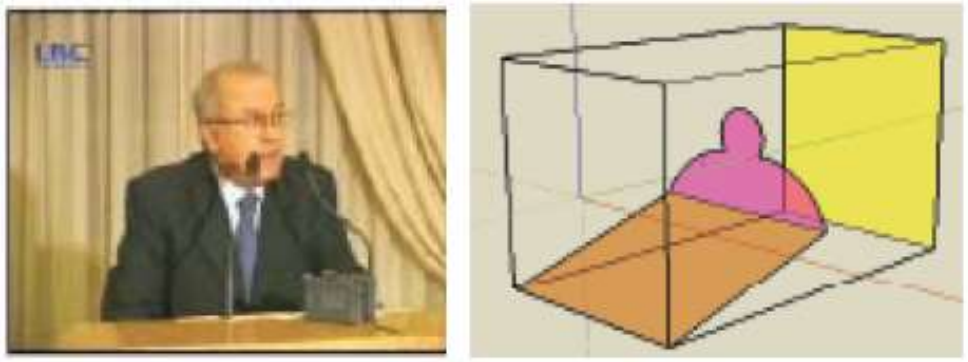
Slide: Derek Hoiem

Wide variety of possible representations

Scene-Level Geometric Description



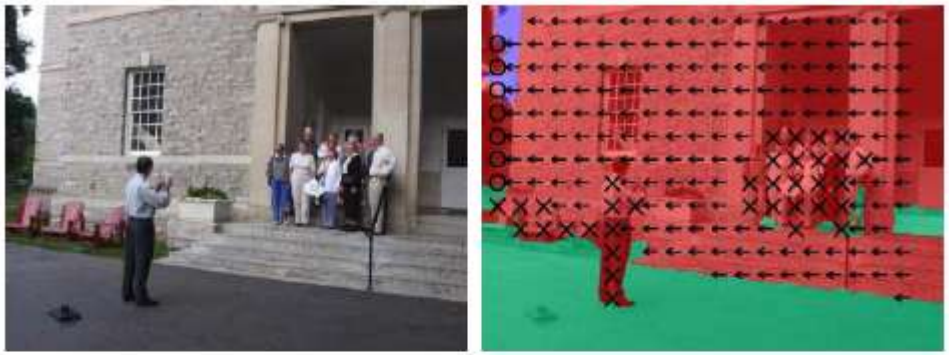
a) Gist, Spatial Envelope



b) Stages

Figs from Hoiem/Savarese Draft

Retinotopic Maps



c) Geometric Context

d) Depth Maps

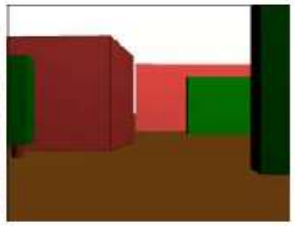
Figs from Hoiem/Savarese Draft

Highly Structured 3D Models

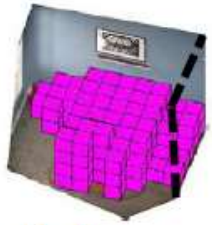
e) Ground Plane

f) Ground Plane with Billboards

g) Ground Plane with Walls



h) Blocks World

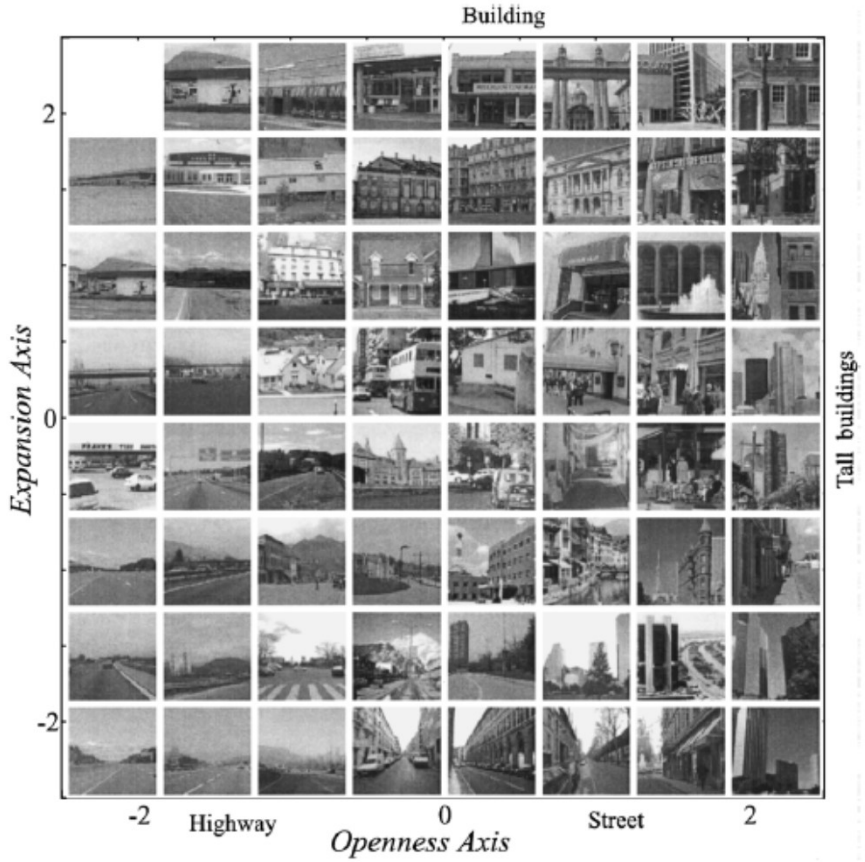


i) 3D Box Model

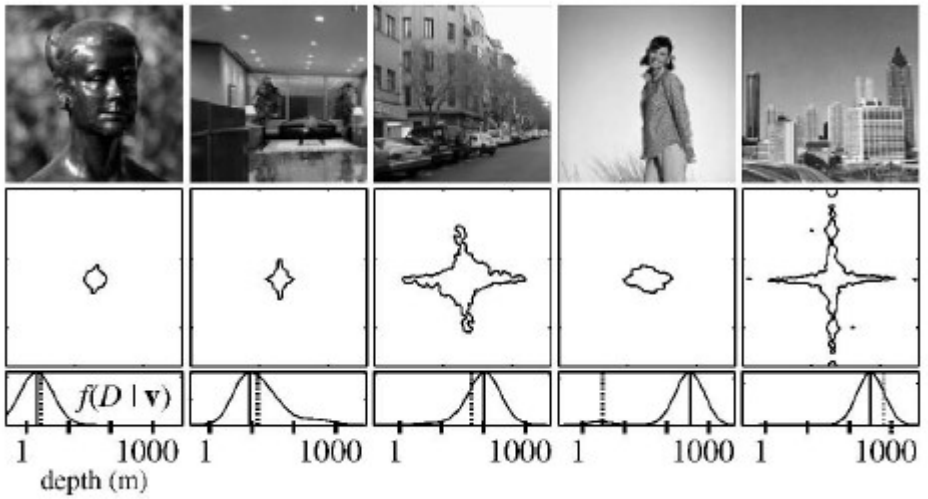
Figs from Hoiem/Savarese Draft

Low detail, Low abstraction

Holistic Scene Space: “Gist”



Oliva & Torralba 2001

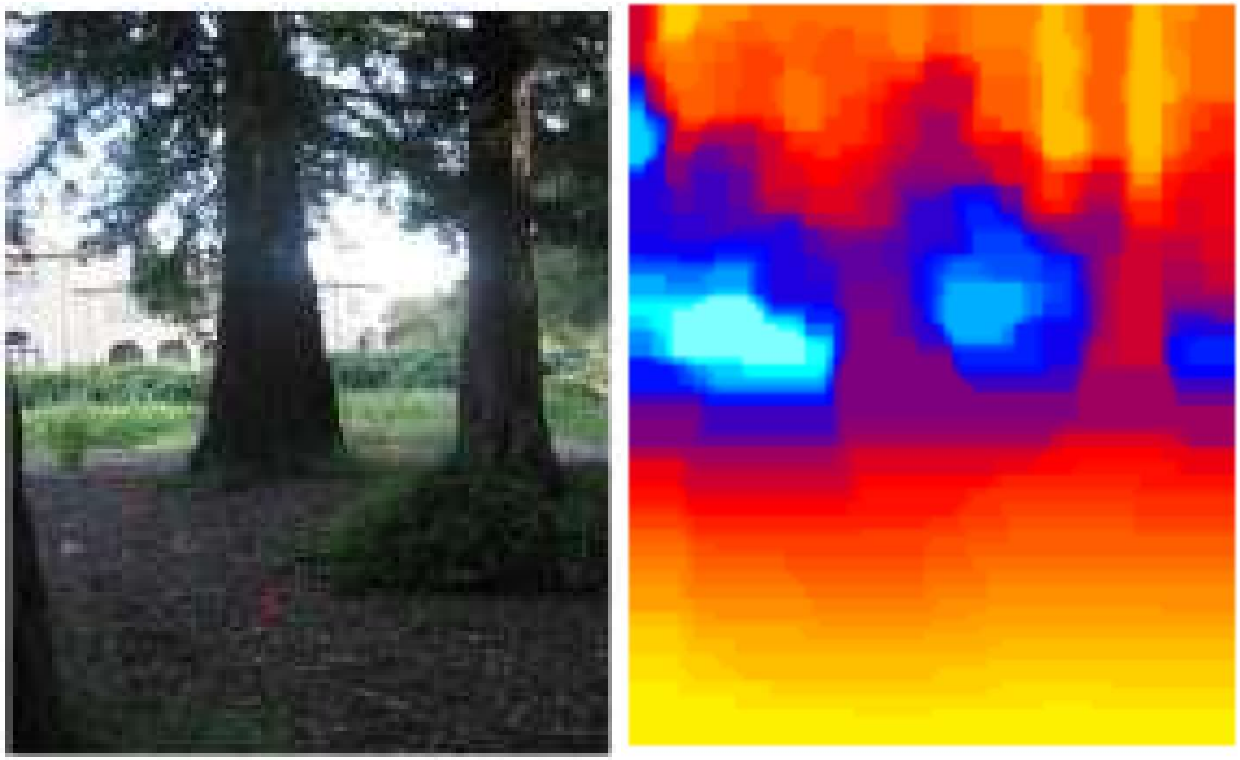


Torralba & Oliva 2002

Slide: Derek Hoiem

High detail, Low abstraction

Depth Map

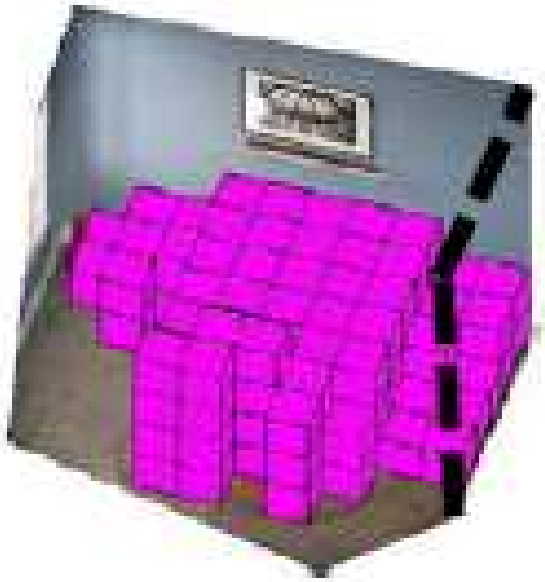


Saxena, Chung & Ng 2005, 2007

Slide: Derek Hoiem

Medium detail, High abstraction

Room as a Box



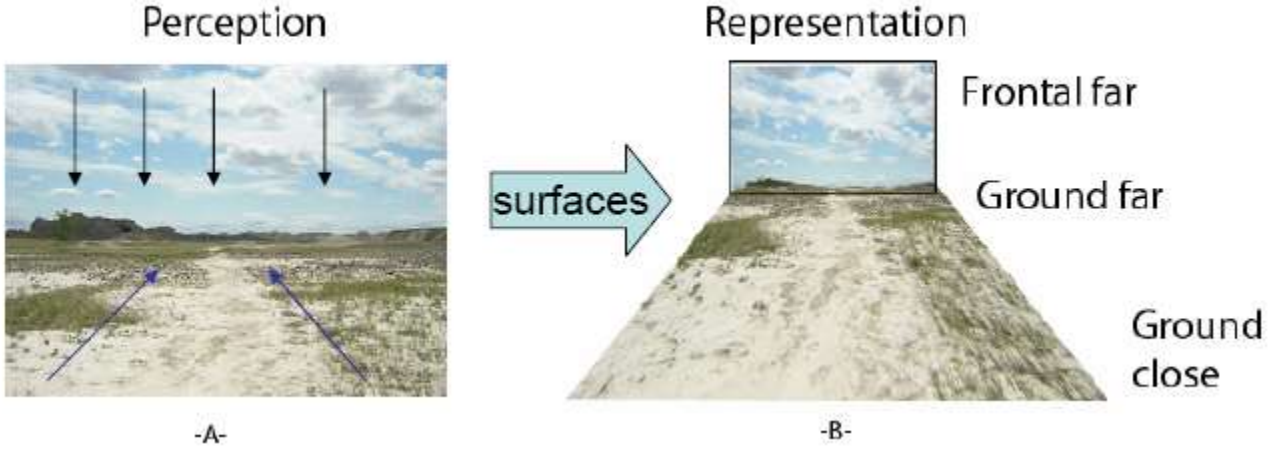
Hedau Hoiem Forsyth 2009

Slide: Derek Hoiem

Gibson's Surface Layout

• Gibson: "The elementary impressions of a visual world are those of surface and edge." *The Perception of the Visual World* (1950)

• Focus on texture gradients



slide from Aude Oliva

Gibson's Surface Layout

Vertical surface texture

Ground surface texture

slide from Aude Oliva

Surface Layout: describe 3D surfaces with geometric classes

Sky

Vertical

Support

Non-Planar Porous

Non-Planar Solid

Planar (Left/Center/Right)

Slide: Derek Hoiem

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

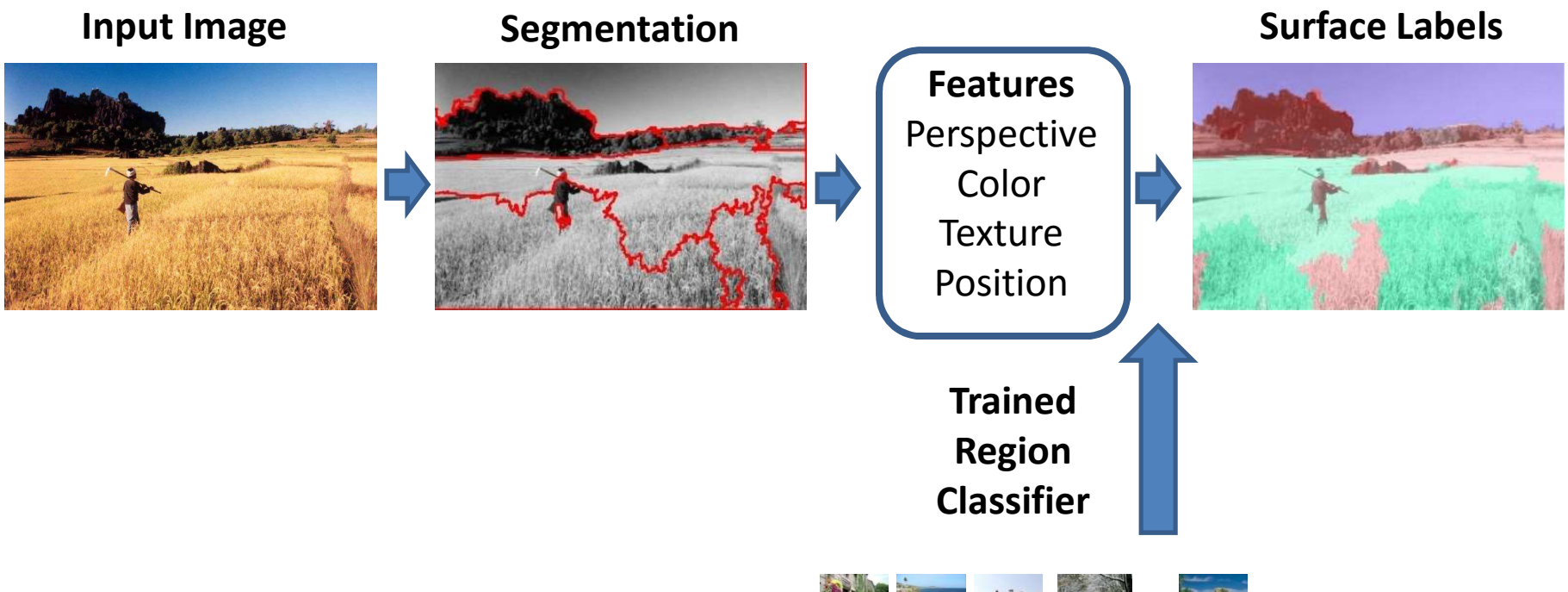
Vanishing points, lines

Color, texture, image location

Texture gradient

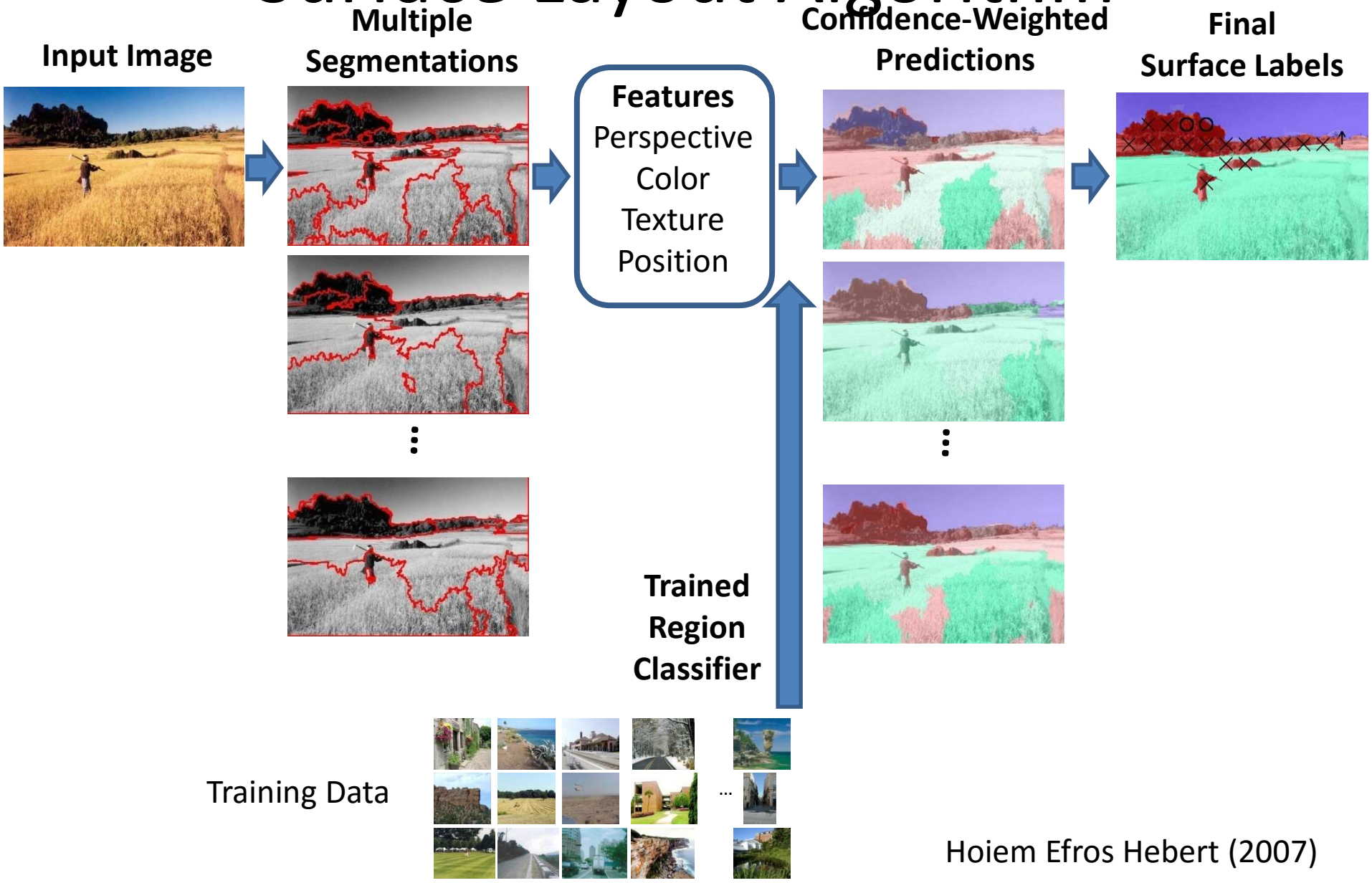
Slide: Derek Hoiem

Surface Layout Algorithm



Hoiem Efros Hebert (2007)

Surface Layout Algorithm



Hoiem Efros Hebert (2007)

Surface Description Result

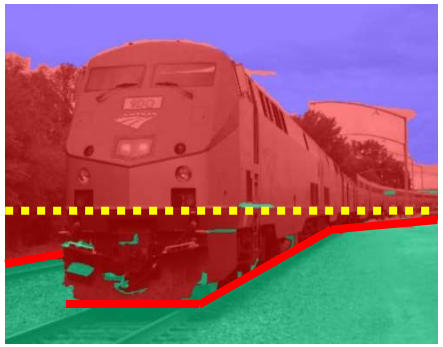
Slide: Derek Hoiem

Automatic Photo Popup

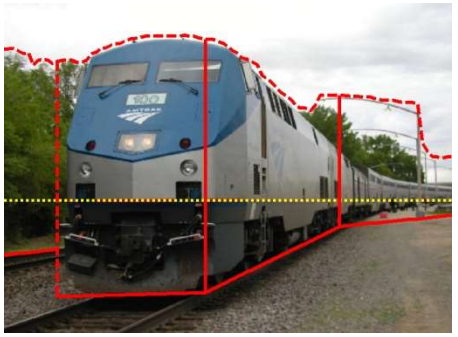
Labeled Image

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

What about more organized but complex spaces?

Other excellent works include:

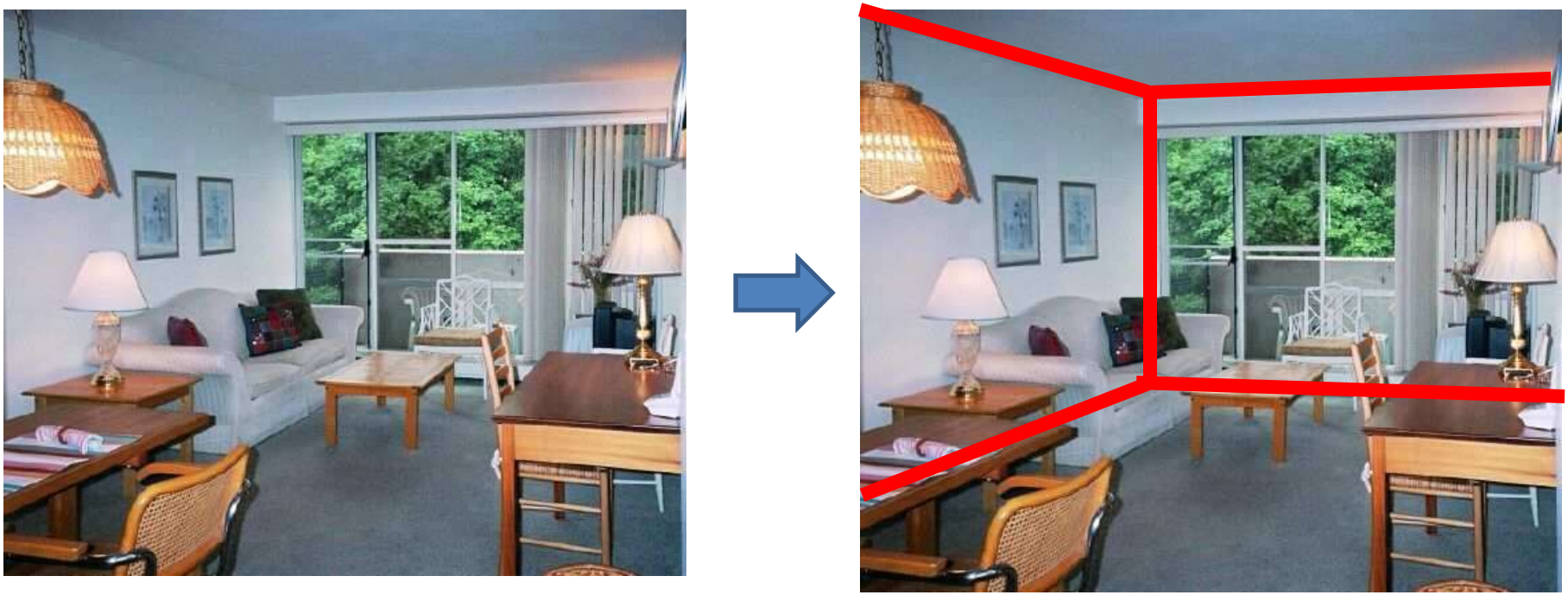
Saxena Sun Ng (2009)

Lee Kanade Hebert (2009)

Gupta Efros Hebert (2010)

Slide: Derek Hoiem

The room as a box



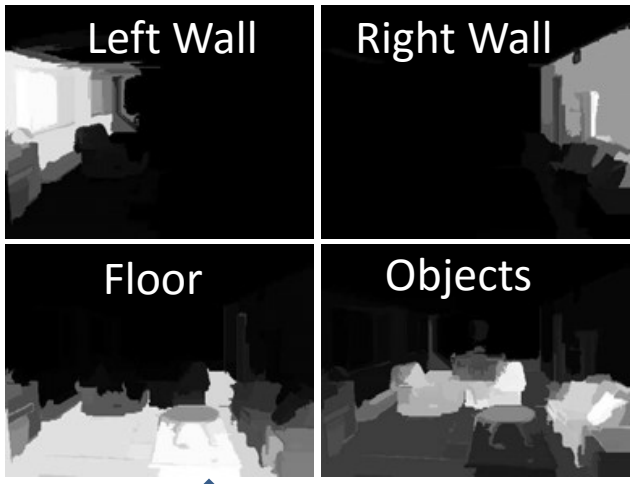
Hedau Hoiem Forsyth (2009)

Recovering the box layout

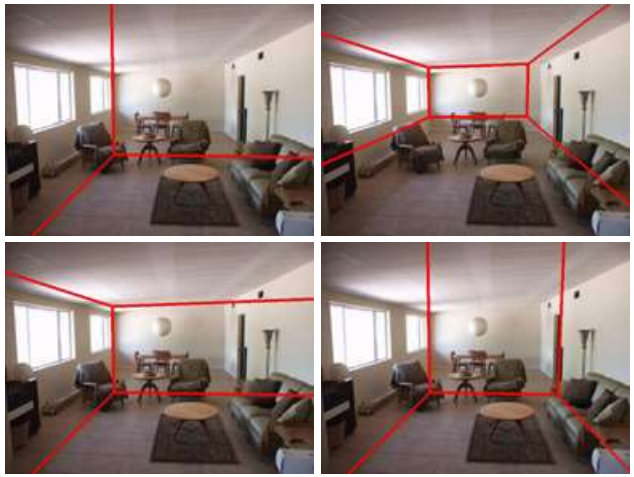
Vertical VP

Detected Edges
+ Vanishing Points

Surface Label Confidences



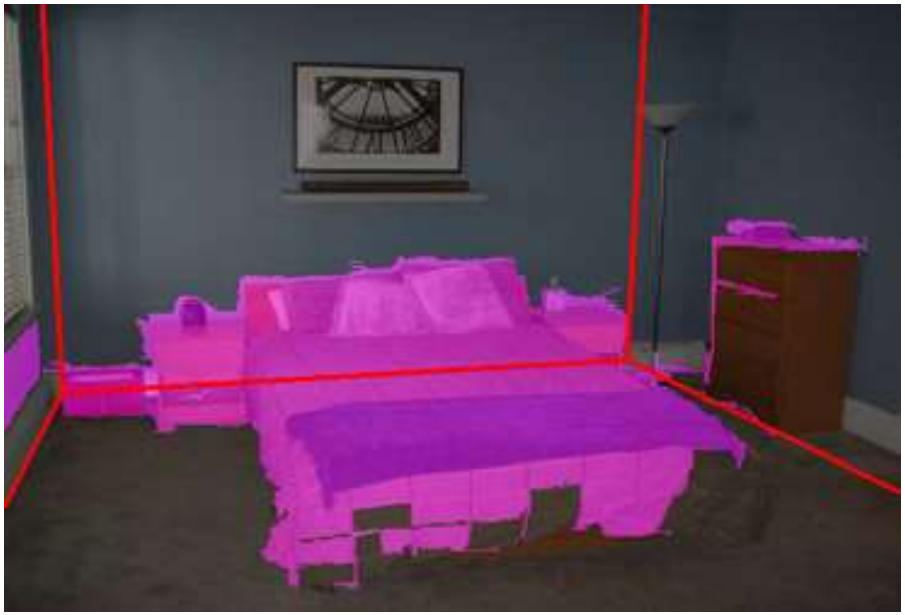
Joint Inference



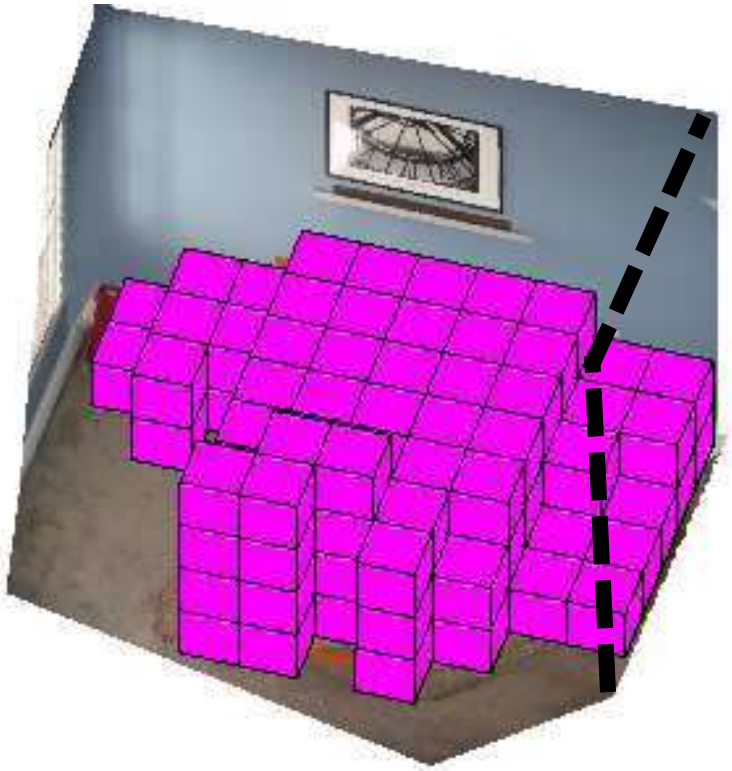
Hypothesized Boxes

Most Likely Geometry

Estimate room's physical space from one image



Estimated "Box" Geometry + Object Pixels

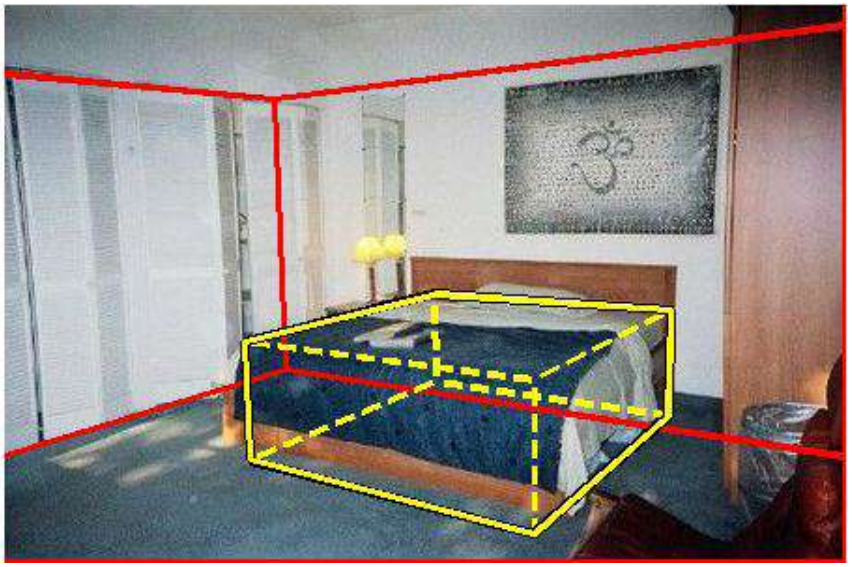


3D Reconstruction + Estimated Occupied Volume

Detecting 3D bed positions in an image

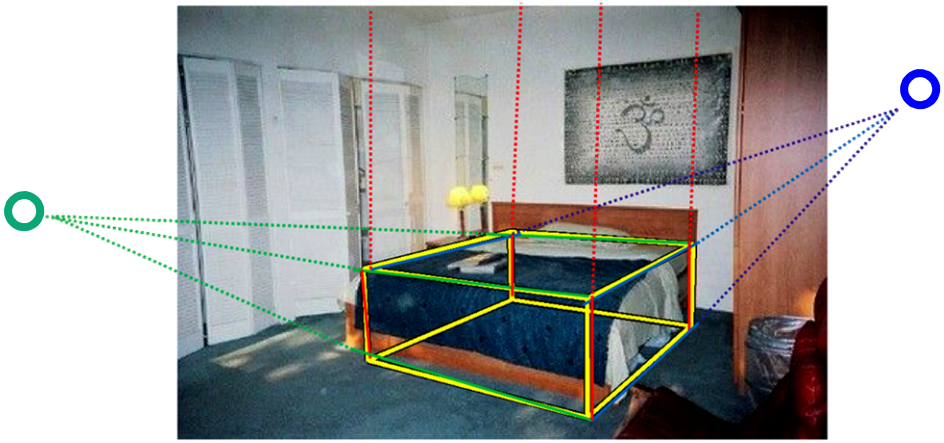
2D Bed Detection

3D Bed Detection with Scene Geometry

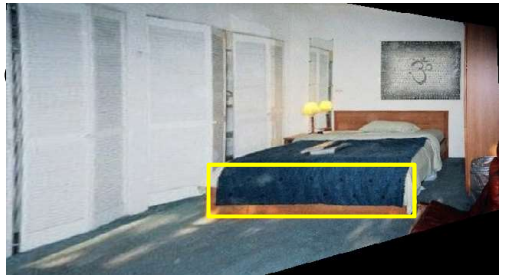


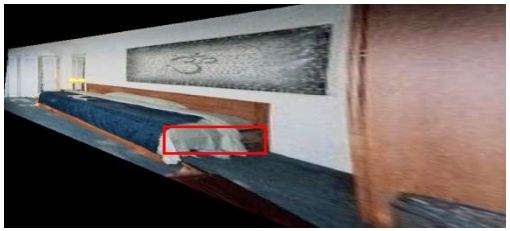
Hedau Hoiem Forsyth (2010)

Searching for beds in room coordin

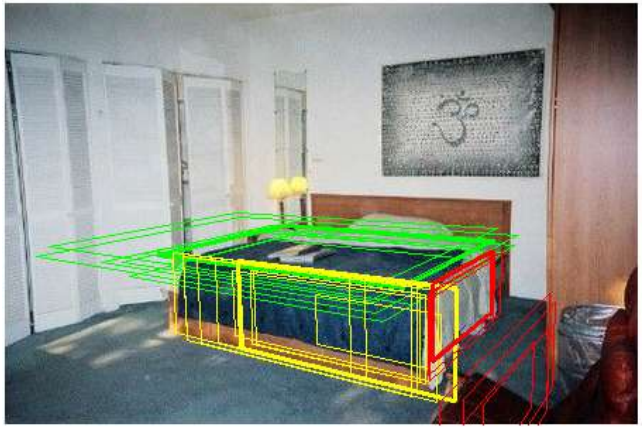


Recover Room Coordinates



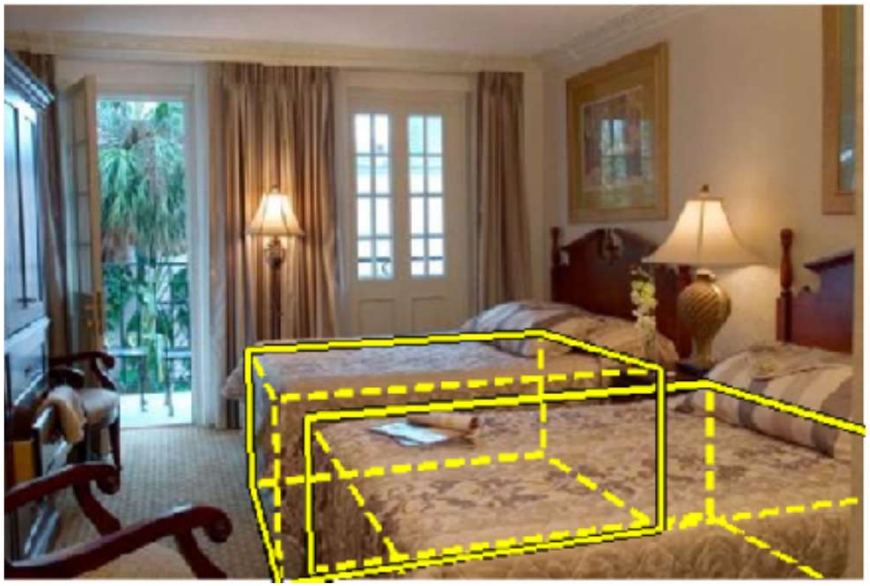


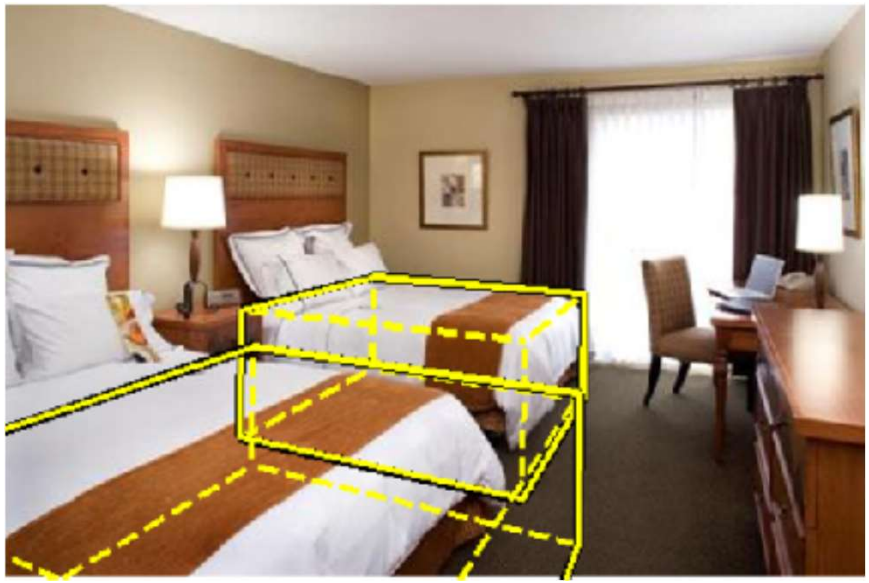
Rectify Features to Room Coordinates



Rectified Sliding Windows

3D bed detection from an image

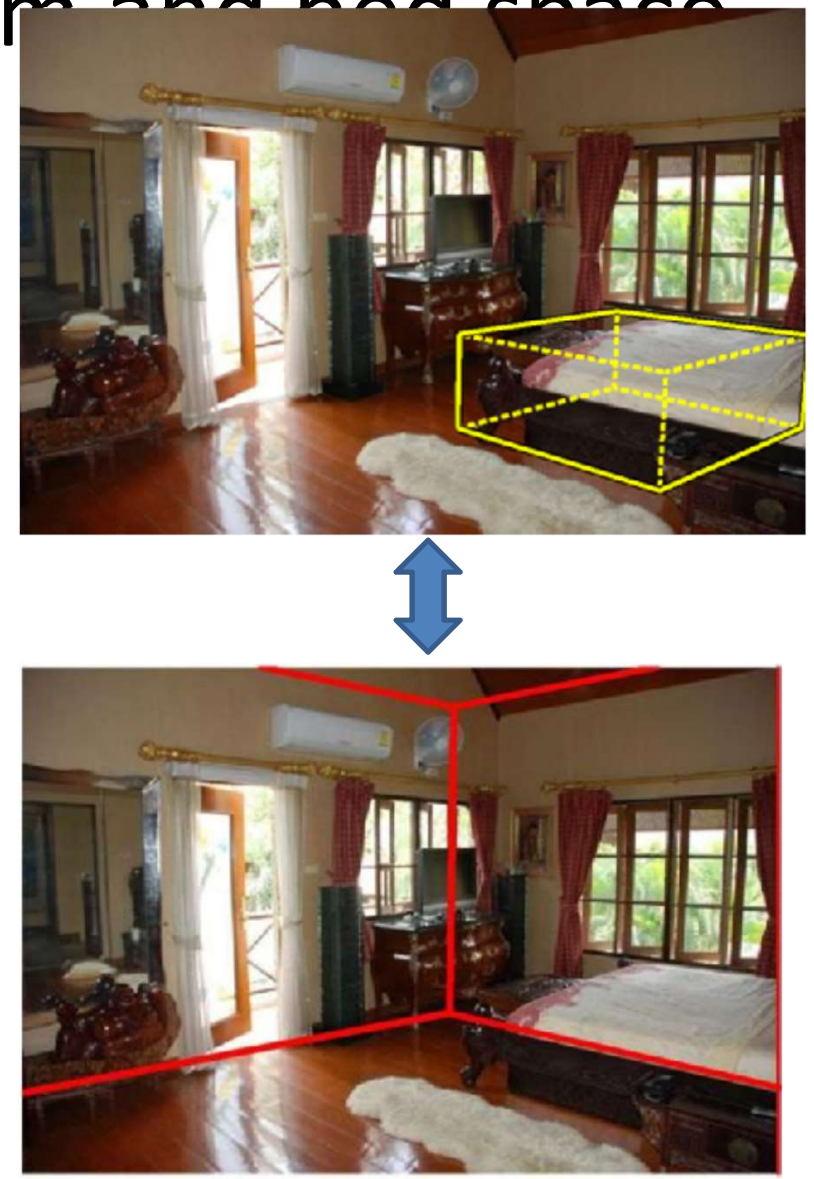




Reason about 3D room and bed space

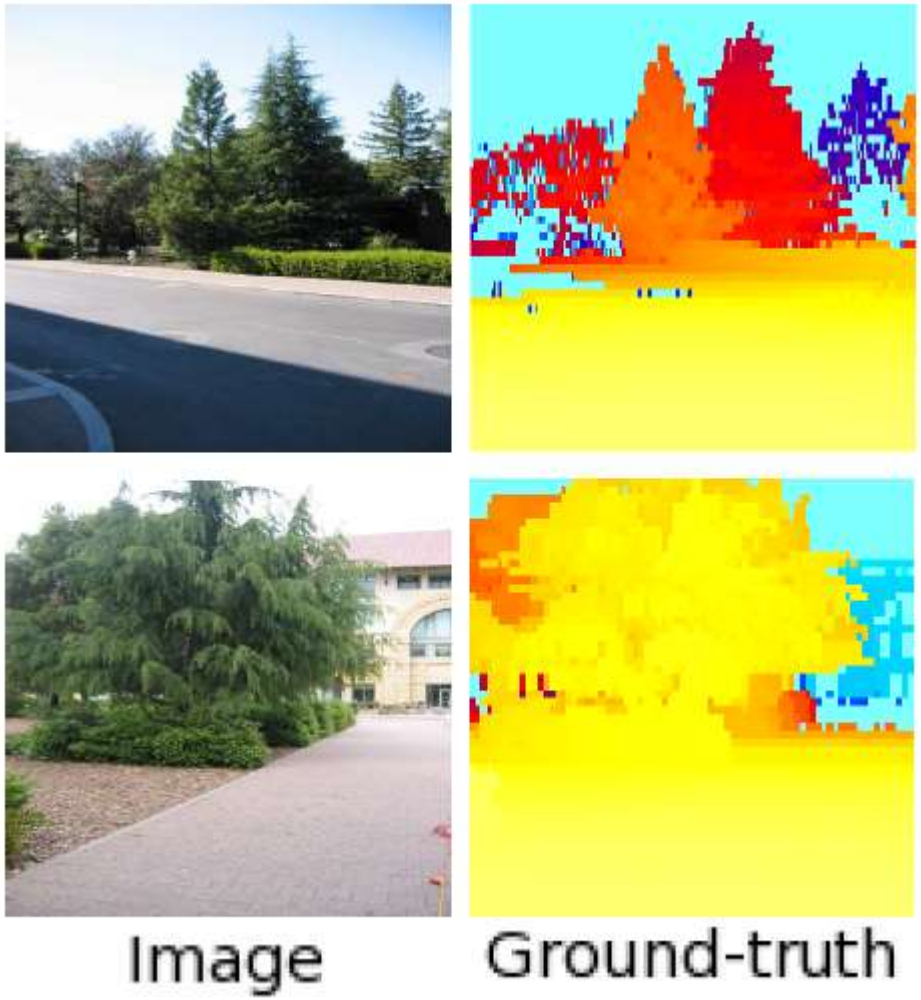
Joint Inference with Priors

- Beds close to walls
- Beds within room
- Consistent bed/wall size
- Two objects cannot occupy the same space



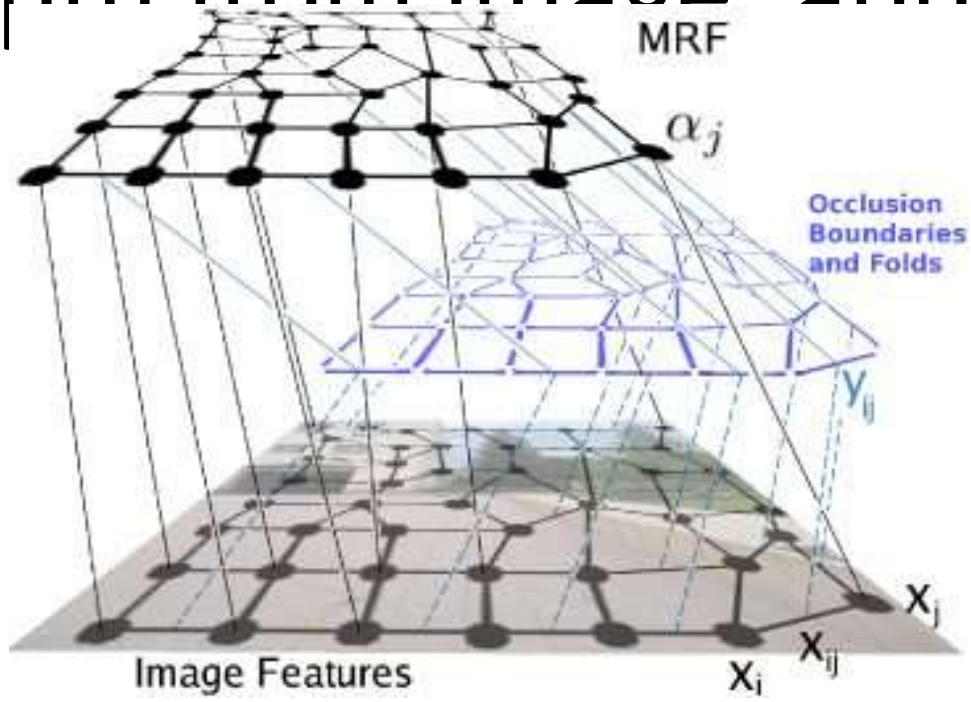
Hedau Hoiem Forsyth (2010)

Depth Estimates from an Image



Saxena et al. 2005, 2008

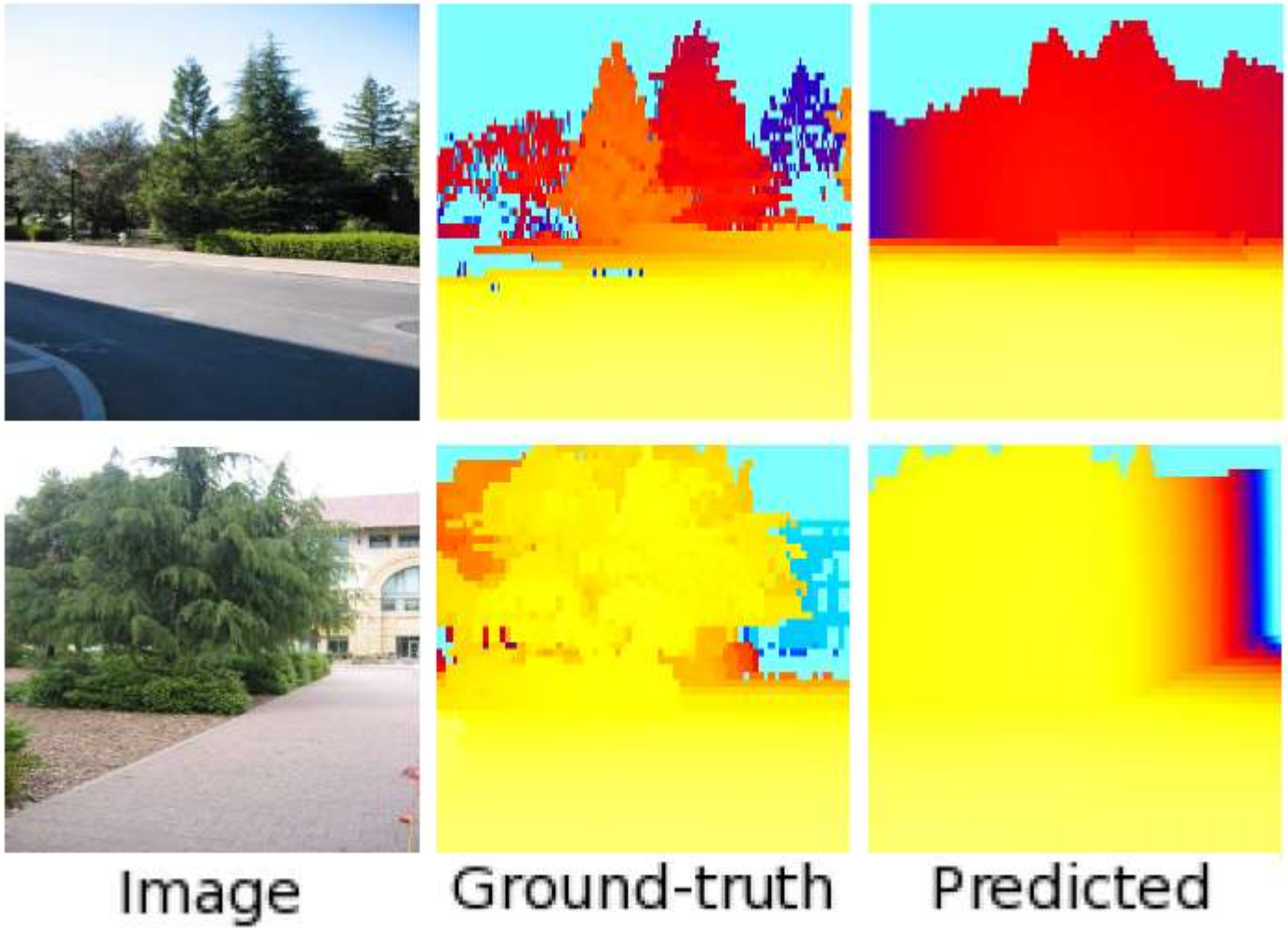
Depth from Image: approach



1. Divide image into superpixels
2. Compute features for each superpixel
   – Position, color, texture, shape
3. Predict 3D plane parameters for each superpixel using features
4. Estimate confidence in prediction using features
5. Global inference, incorporating constraints of connected structure, co-planarity, co-linearity

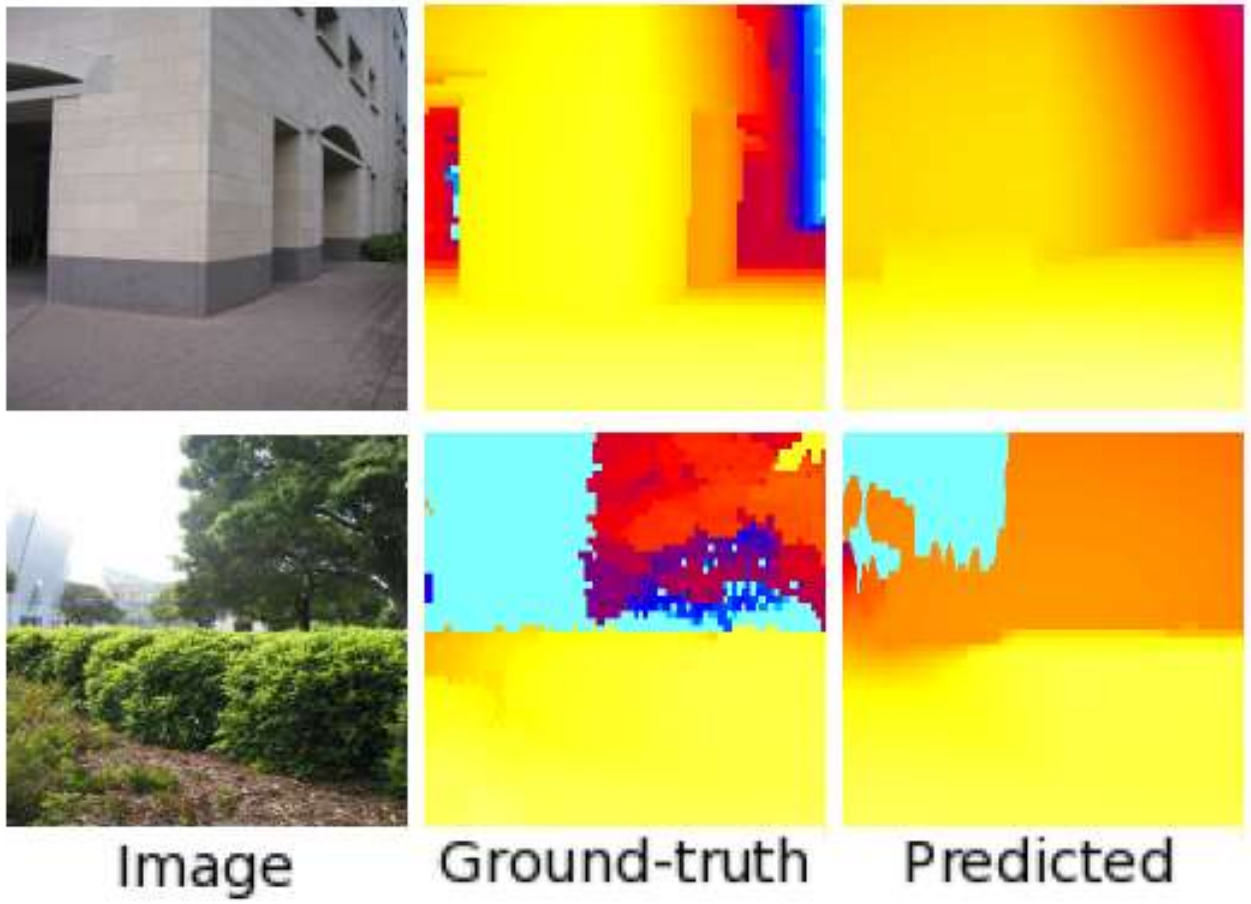
Saxena et al. 2008

Depth Estimates from an Image



Saxena et al. 2005, 2008

Depth Estimates from an Image



Saxena et al. 2008

Depth from Image: Reconstructions

Input

Novel View

Saxena et al. 2008

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