- Project 4
- MLDemos - http://mldemos.epfl.ch/
- Spring 2012: CS 2951-B, Data-Driven Vision and Graphics, 11-12, CIT 345

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.

Context provides clues for function

- What is this?

These examples from Antonio Torralba

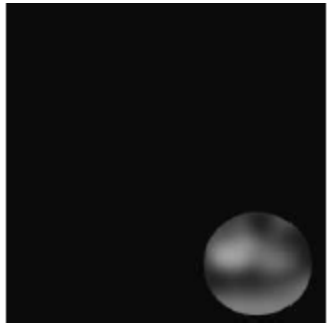
Context provides clues for function

- What is this?

- Now can you tell?

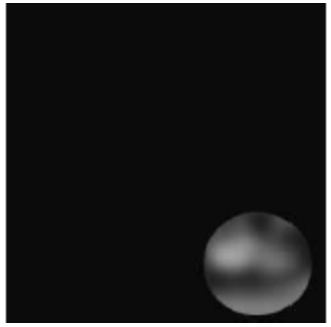
Sometimes context is *the* major component of recognition

- What is this?



Sometimes context is *the* major component of recognition

- What is this?



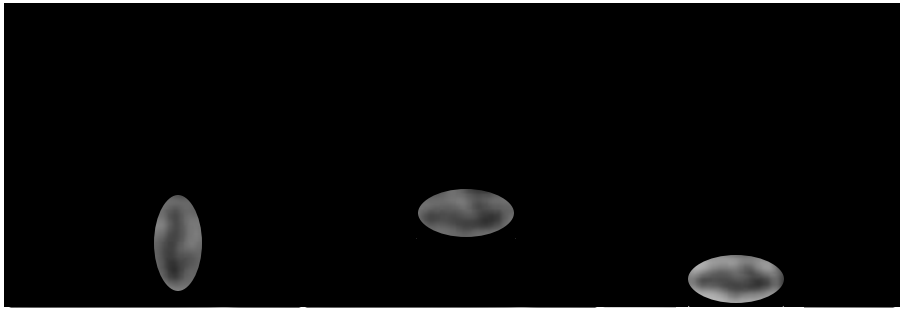
- Now can you tell?

More Low-Res

- What are these blobs?

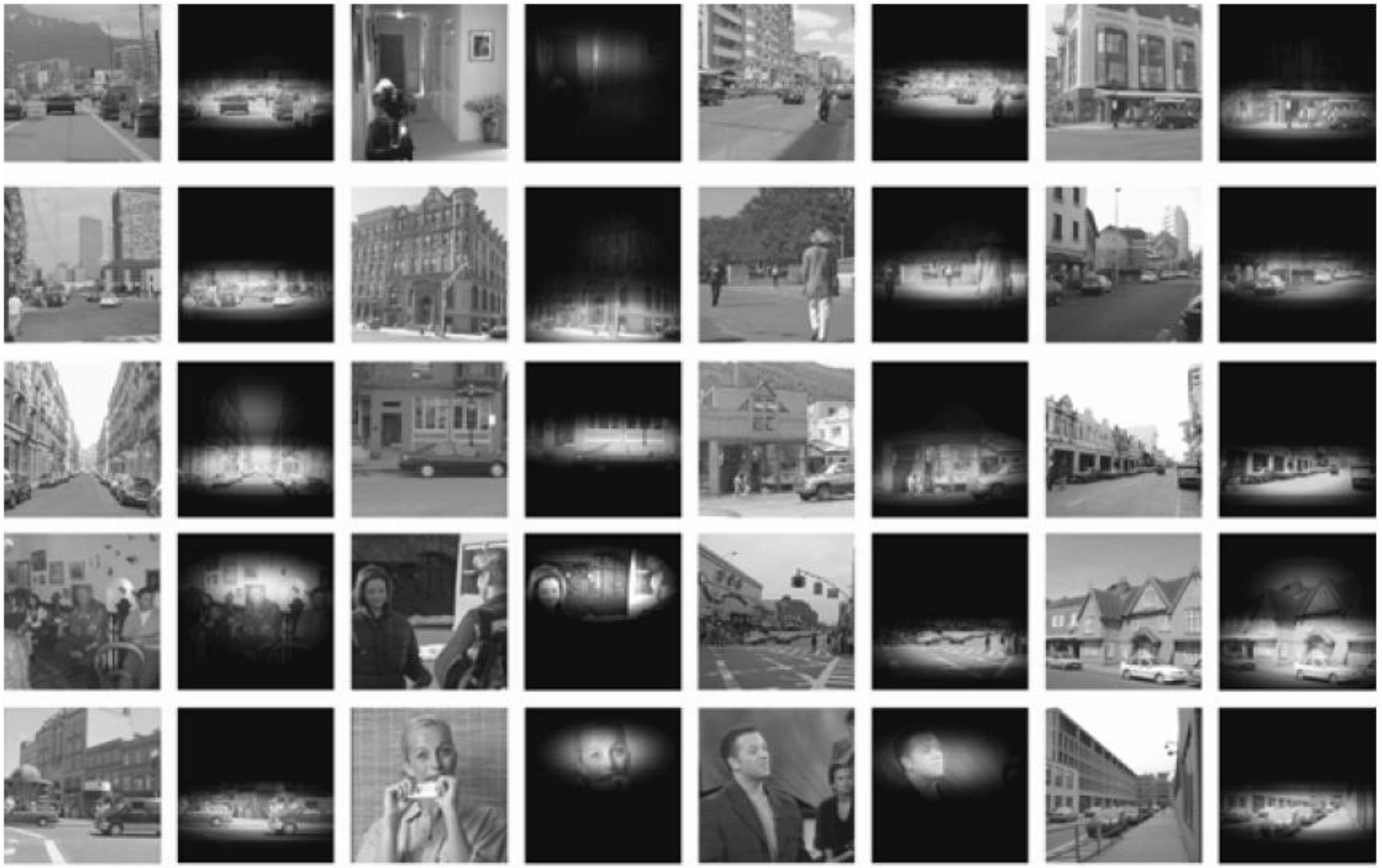
More Low-Res

- The same pixels! (a car)



The Context Challenge

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



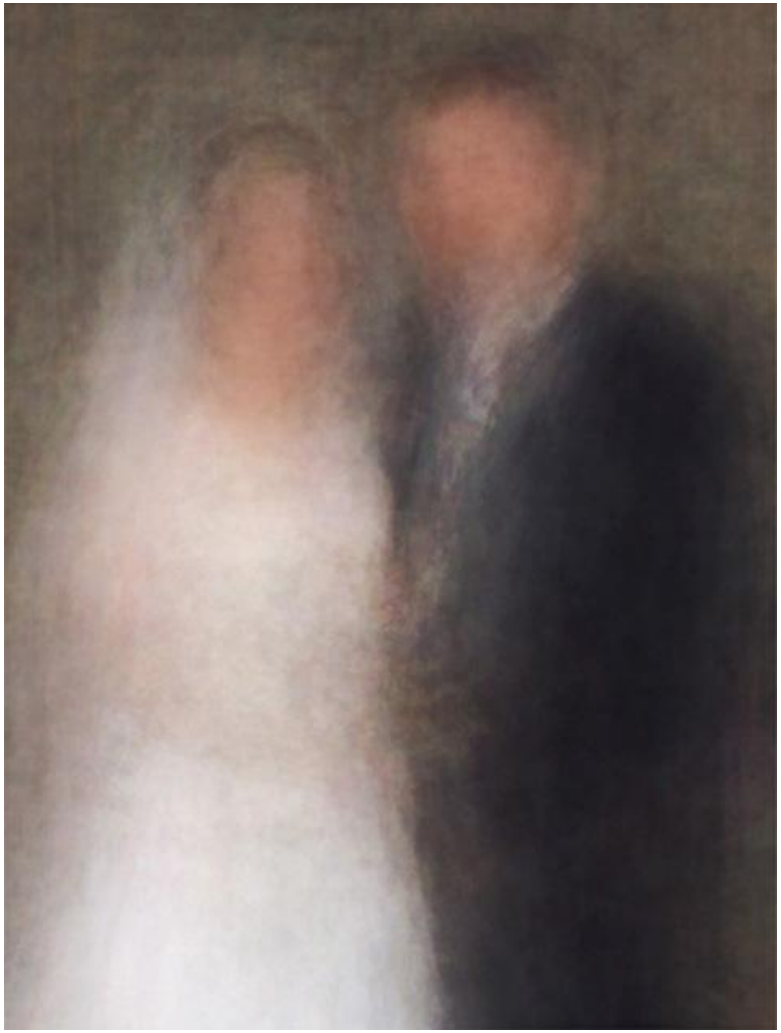
No local face detector! Just context from Scene Statistics

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?

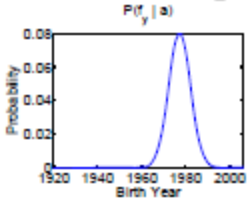
Andrew Gallagher: http://chenlab.ece.cornell.edu/people/Andy/projectpage_names.html

Cultural context

Age given Appearance

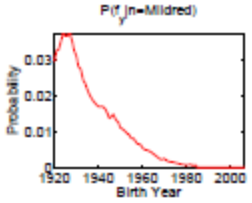
Age given Name

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



Mildred

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

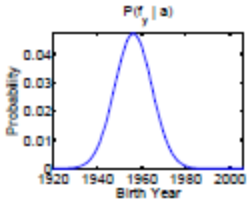


3.88

3.88

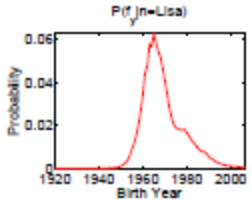
4.77

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



Lisa

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



6.70

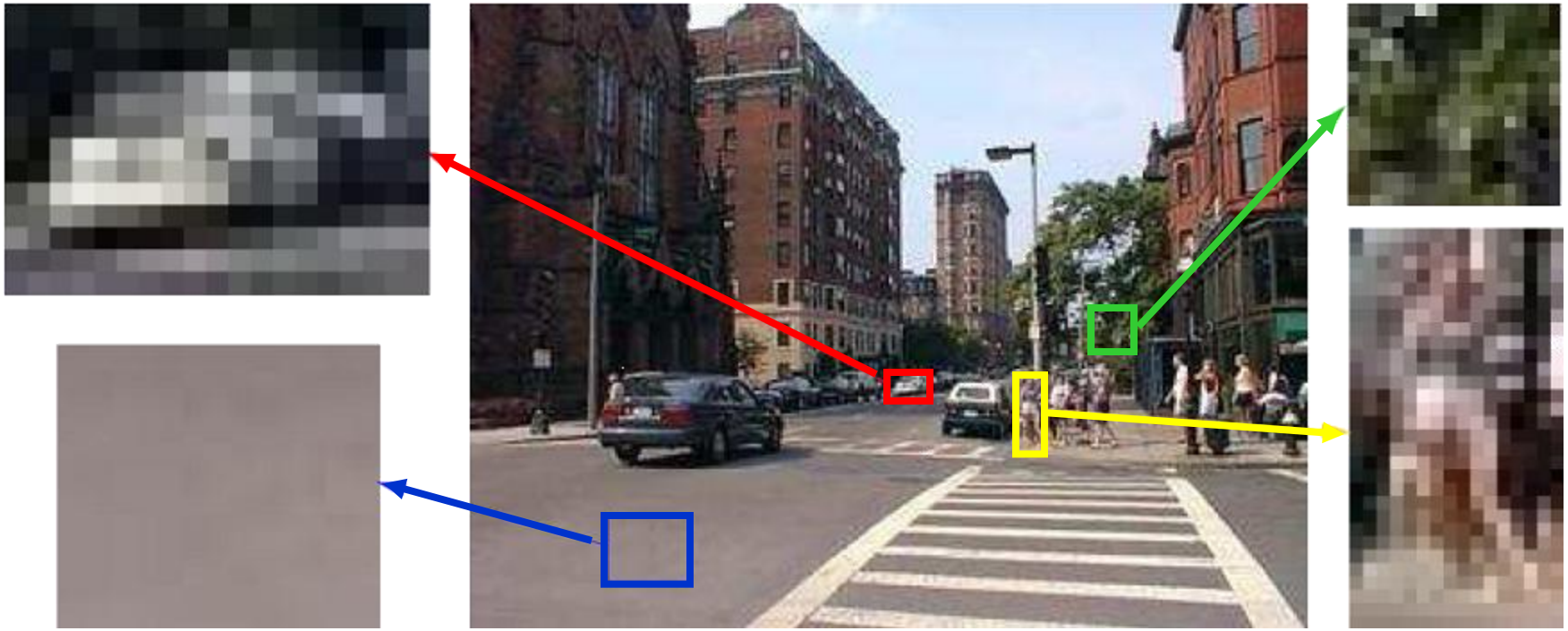
Andrew Gallagher: http://chenlab.ece.cornell.edu/people/Andy/projectpage_names.html

Spatial layout is especially important

1. Context for recognition

Spatial layout is especially important

1. Context for recognition



Spatial layout is especially important

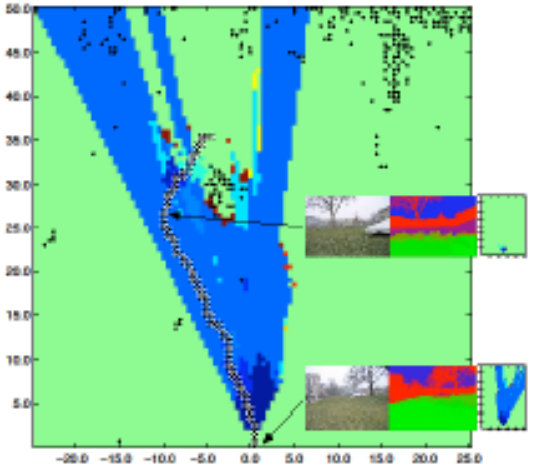
1. Context for recognition
2. Scene understanding

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

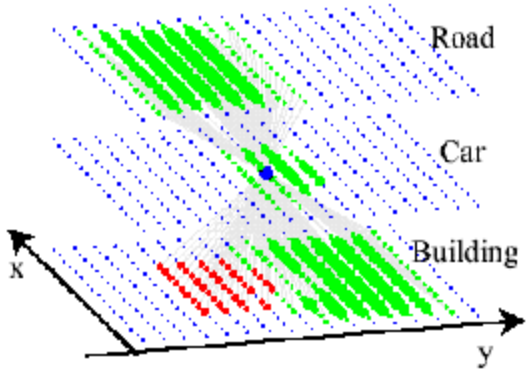


Robot Navigation: Path Planning

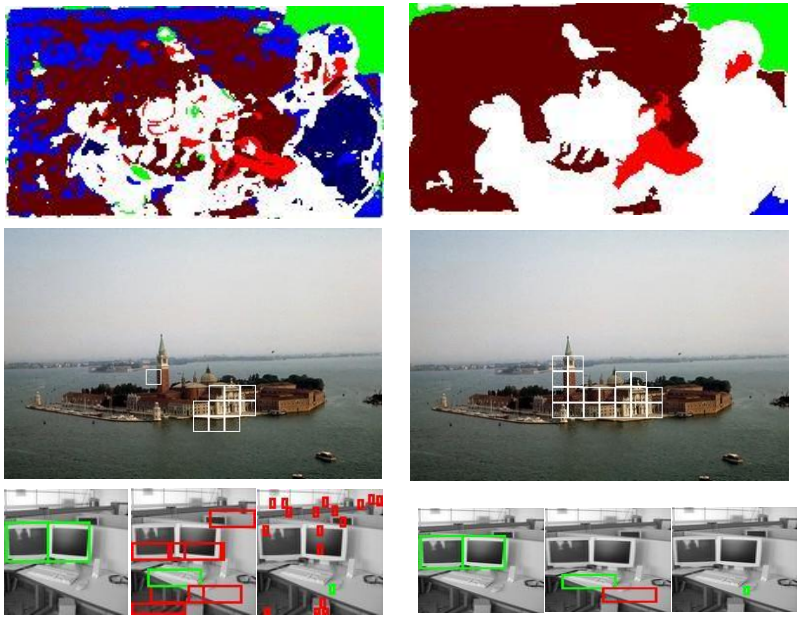
Spatial Layout: 2D vs. 3D

Slide: Derek Hoiem

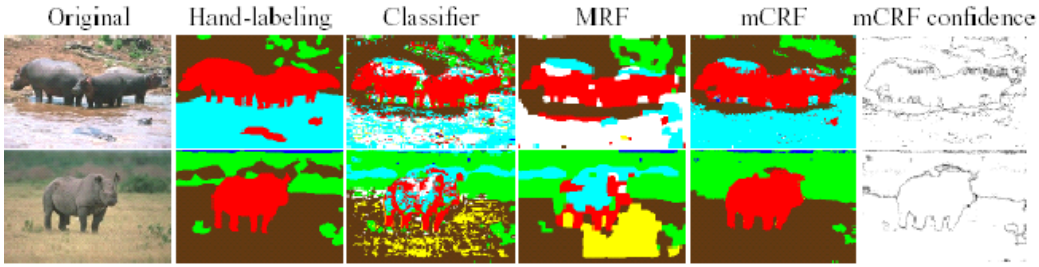
Context in Image Space



[Torralba Murphy Freeman 2004]



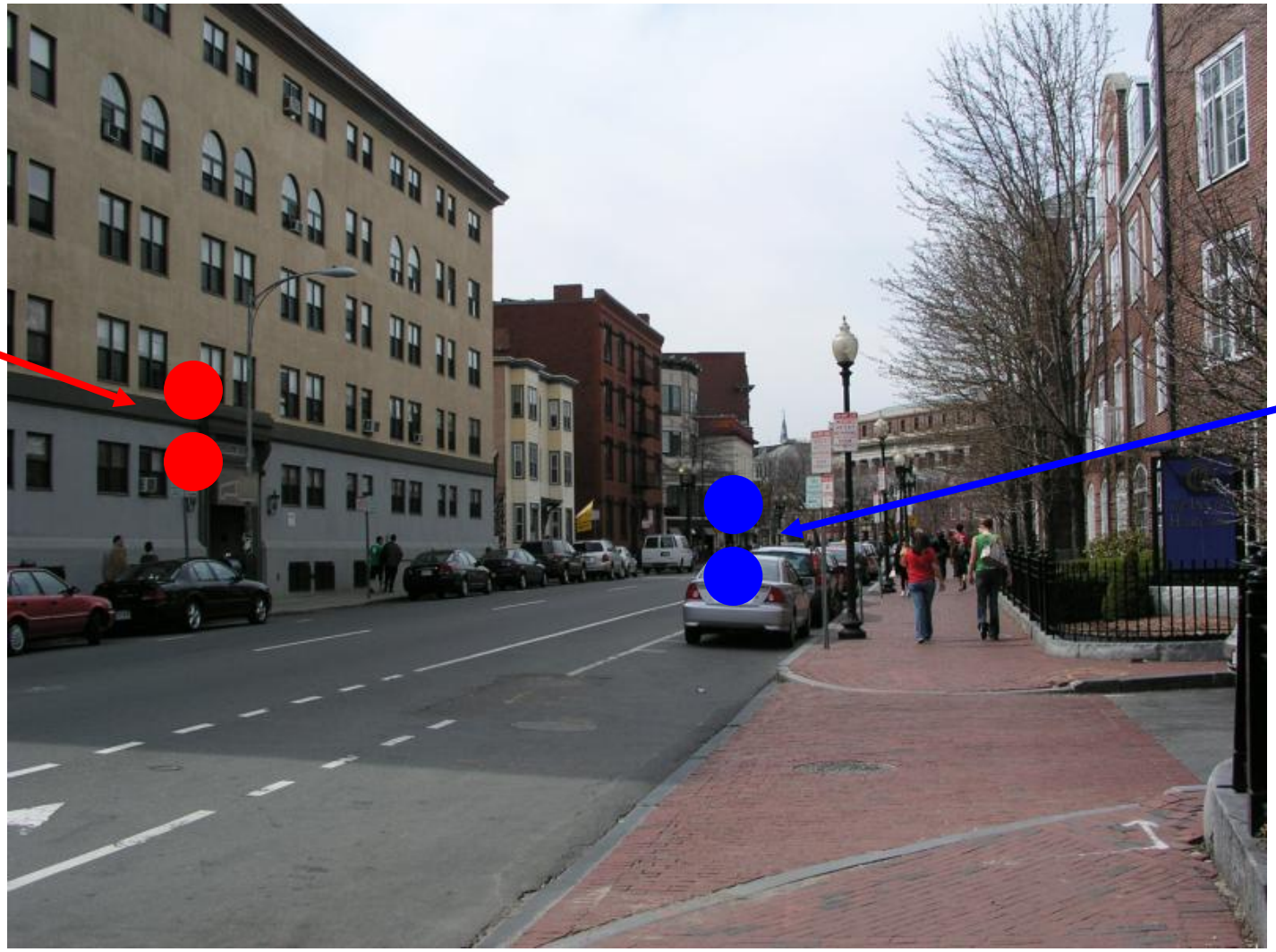
[Kumar Hebert 2005]



[He Zemel Cerreira-Perpiñán 2004]

Slide: Derek Hoiem

But object relations are in 3D…

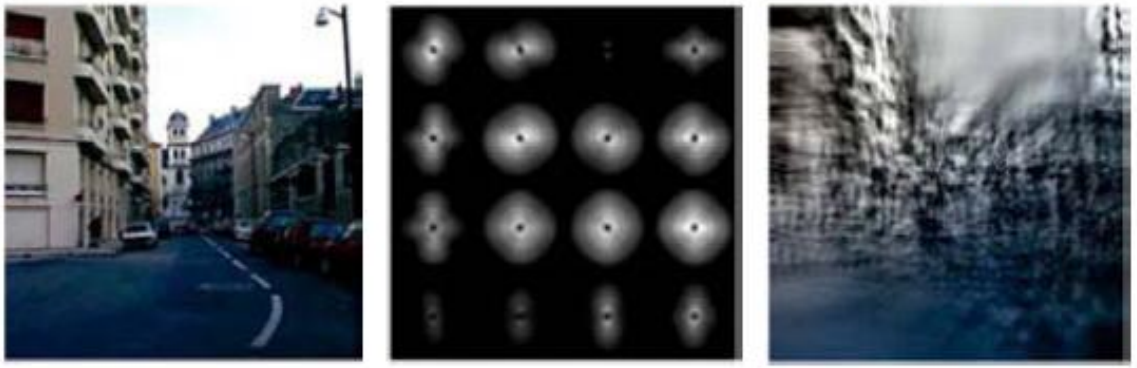


Slide: Derek Hoiem

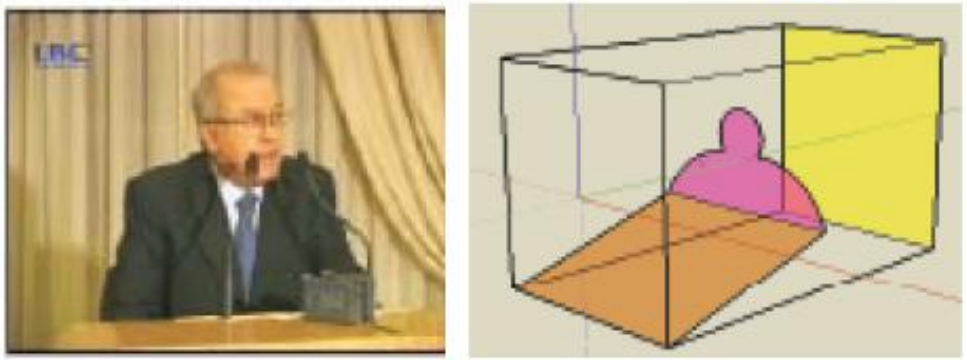
How to represent scene space?

Wide variety of possible representations

Scene-Level Geometric Description



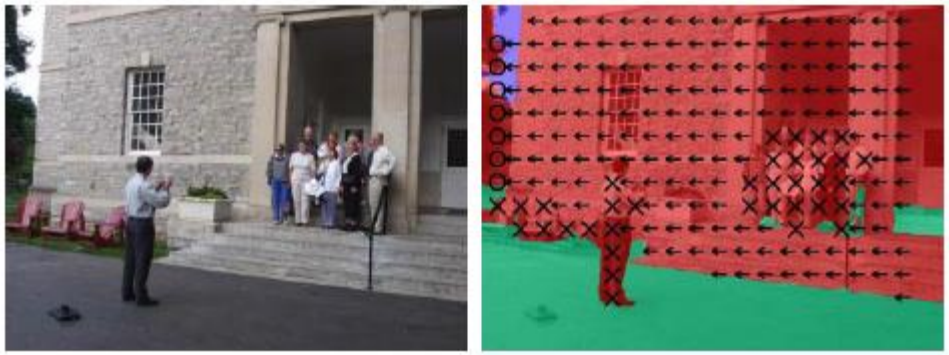
a) Gist, Spatial Envelope



b) Stages

Figs from Hoiem/Savarese Draft

Retinotopic Maps



c) Geometric Context

d) Depth Maps

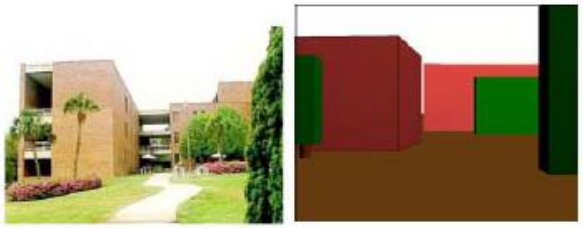
Figs from Hoiem/Savarese Draft

Highly Structured 3D Models

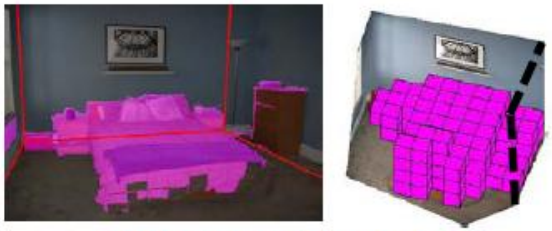
e) Ground Plane

f) Ground Plane with Billboards

g) Ground Plane with Walls



h) Blocks World



i) 3D Box Model

Figs from Hoiem/Savarese Draft

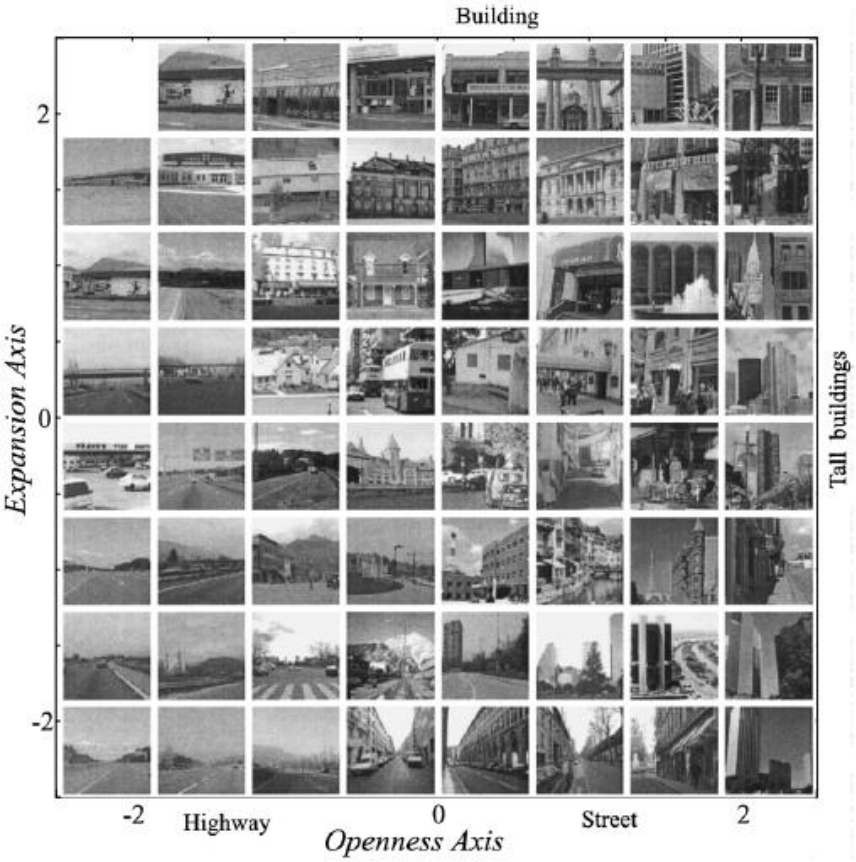
Key Trade-offs

- 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

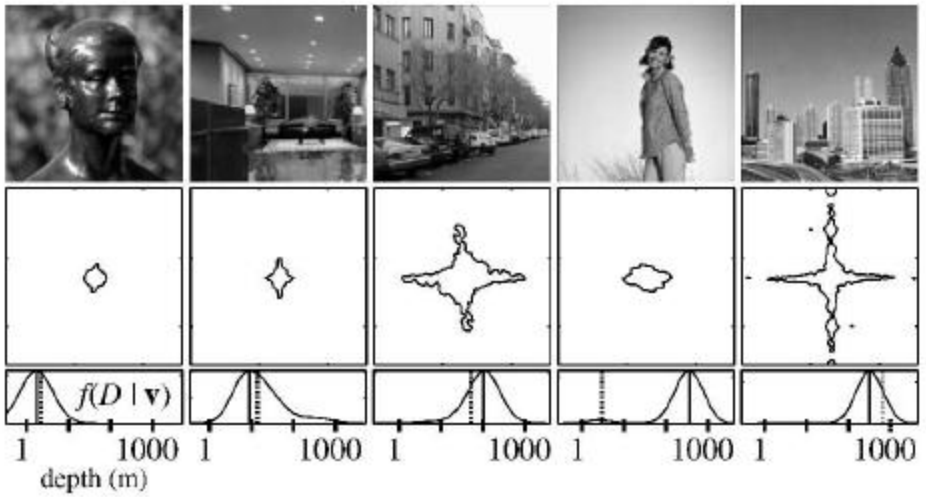
Slide: Derek Hoiem

Low detail, Low abstraction

Holistic Scene Space: “Gist”



Oliva & Torralba 2001

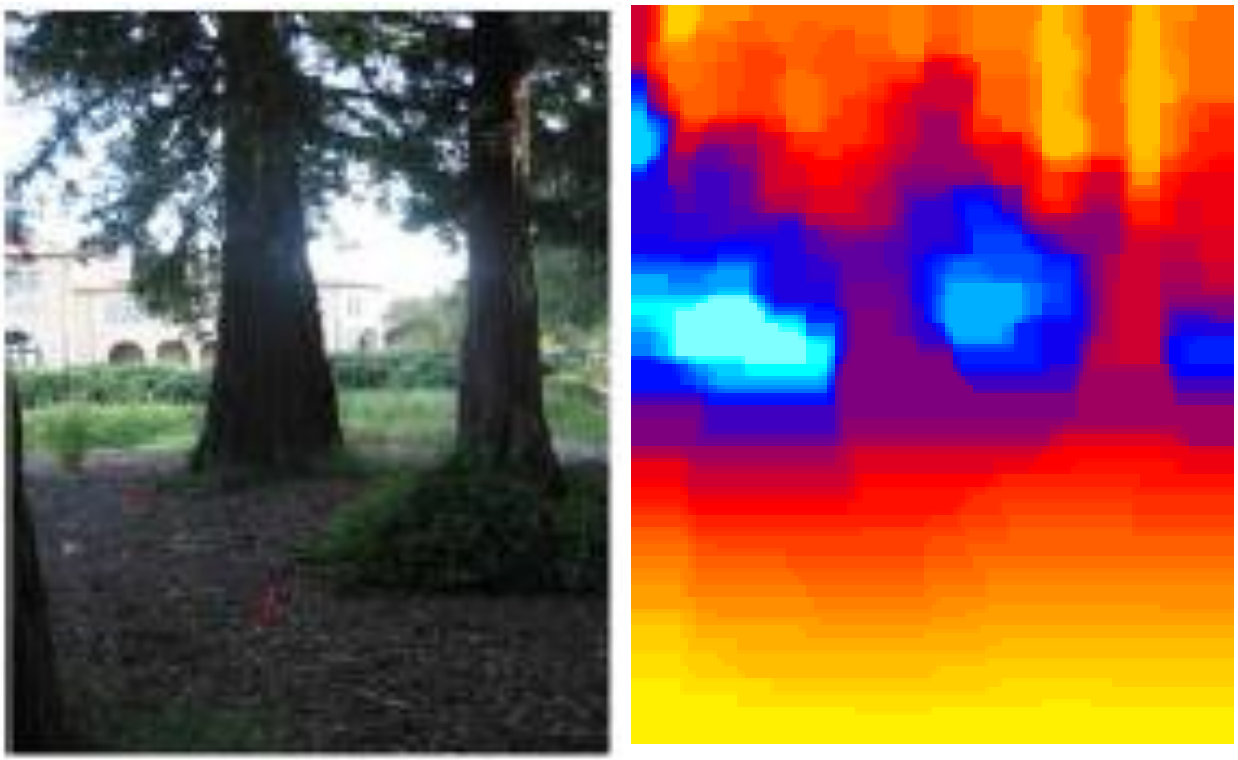


Torralba & Oliva 2002

Slide: Derek Hoiem

High detail, Low abstraction

Depth Map

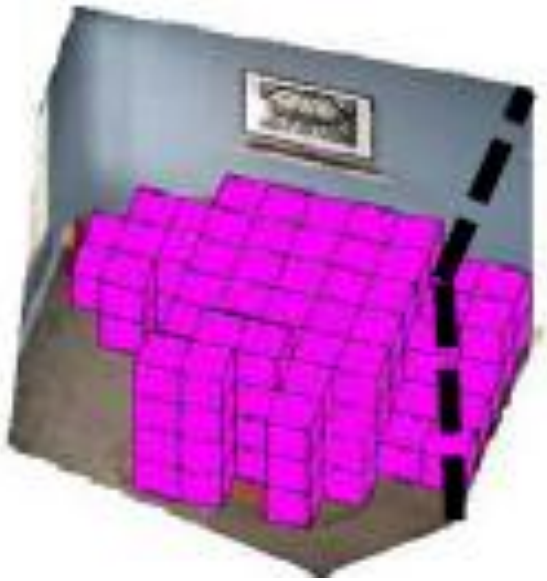


Saxena, Chung & Ng 2005, 2007

Slide: Derek Hoiem

Medium detail, High abstraction

Room as a Box



Hedau Hoiem Forsyth 2009

Slide: Derek Hoiem

A few examples of spatial layout estimation

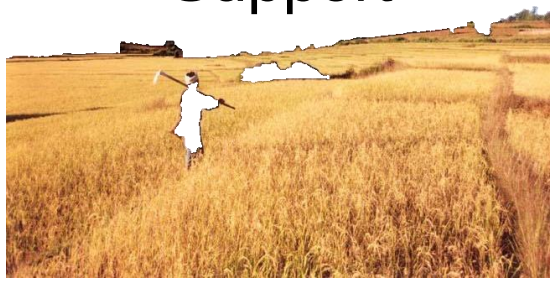
- Surface layout
- The room as a box
- Depth estimation

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

Surface Layout Algorithm

Input Image

Segmentation

Features
Perspective
Color
Texture
Position

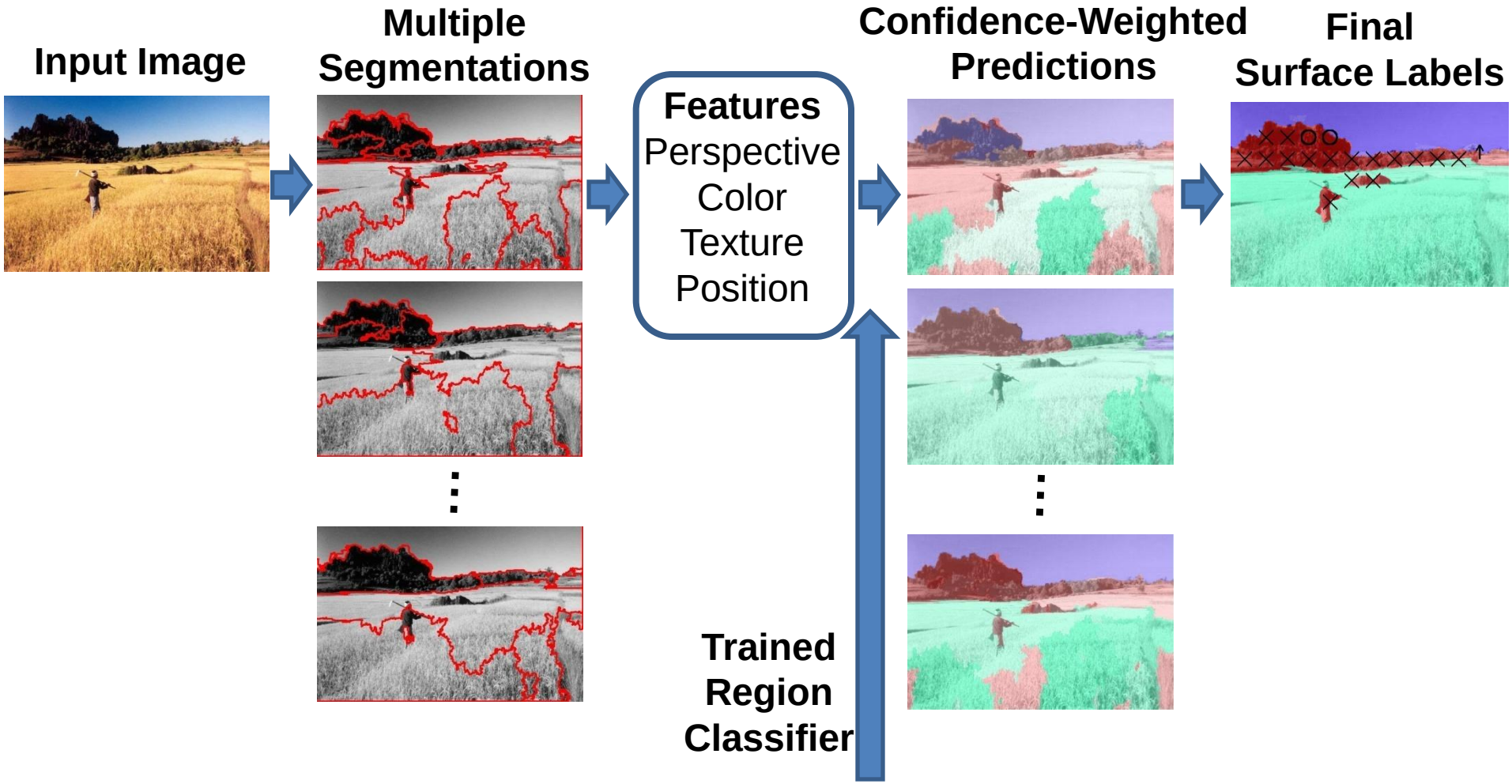
Surface Labels

Trained Region Classifier

Training Data

Hoiem Efros Hebert (2007)

Surface Layout Algorithm



Hoiem Efros Hebert (2007)

Surface Description Result

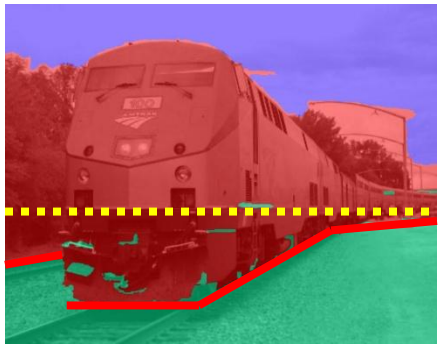
Slide: Derek Hoiem

Automatic Photo Popup

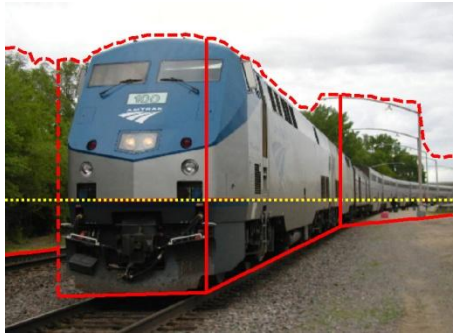
Labeled Image

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

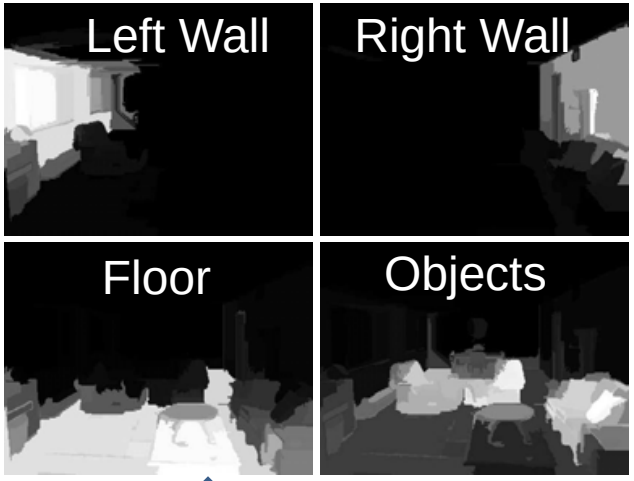
Hedau Hoiem Forsyth (2009)

Recovering the box layout

Vertical VP

Detected Edges + Vanishing Points

Surface Label Confidences

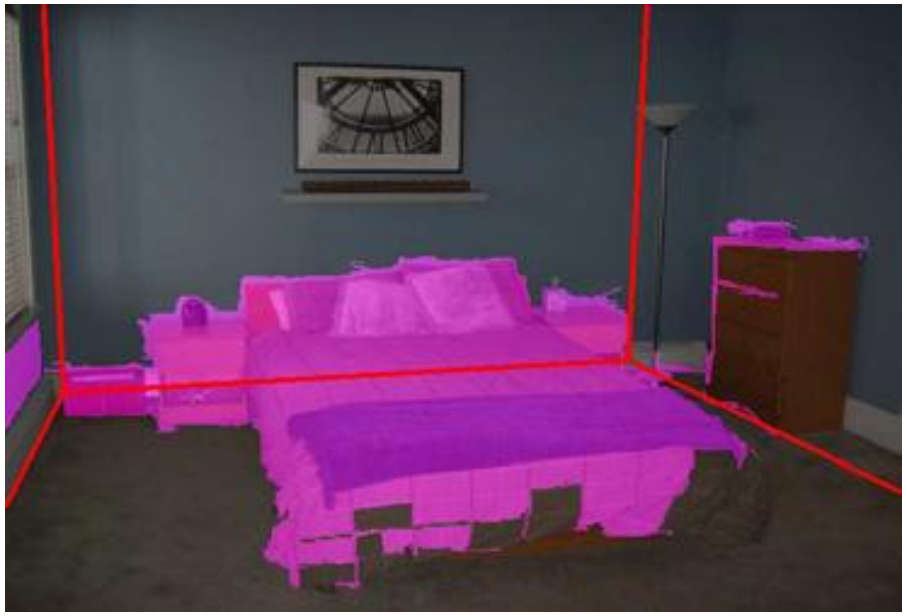


Joint Inference

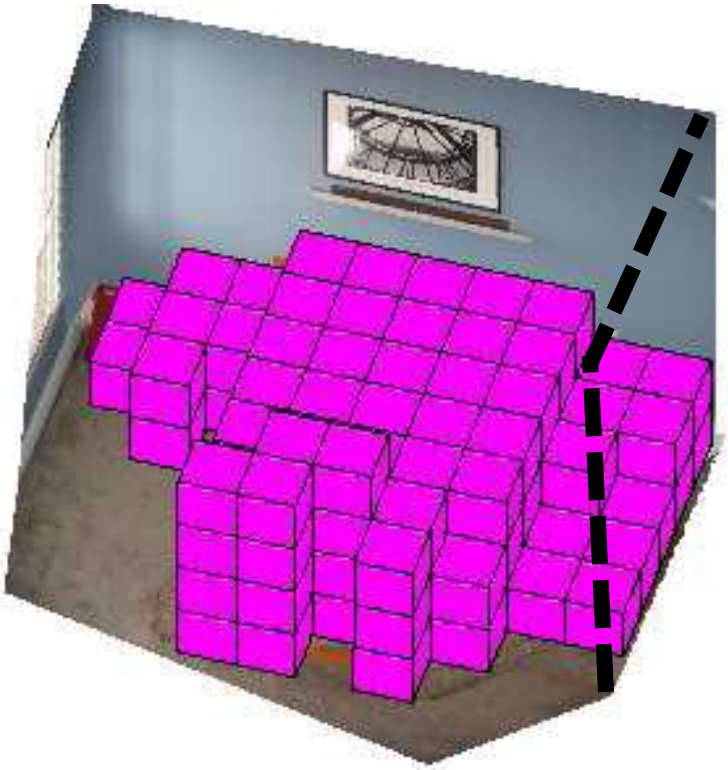
Hypothesized Boxes

Most Likely Geometry

Estimate room's physical space from one image



Estimated "Box" Geometry
+ Object Pixels

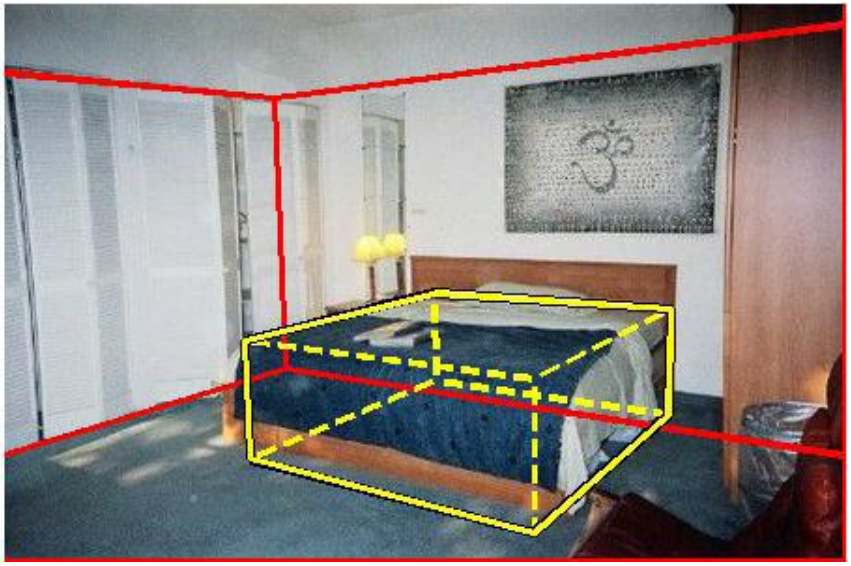


3D Reconstruction +
Estimated Occupied Volume

Detecting 3D bed positions in an image

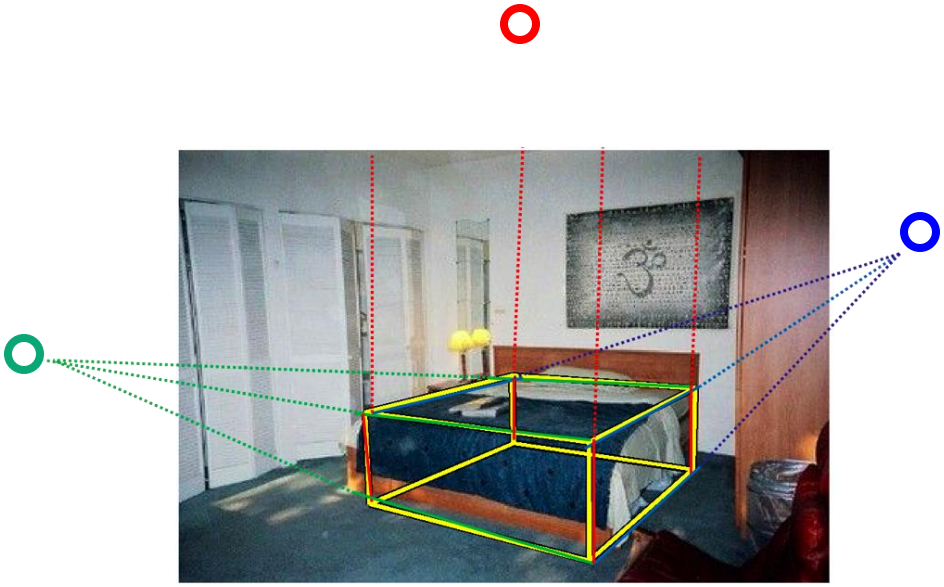
2D Bed Detection

3D Bed Detection with Scene Geometry

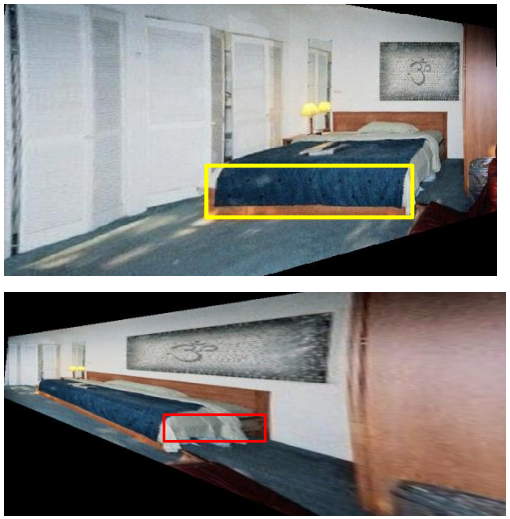


Hedau Hoiem Forsyth (2010)

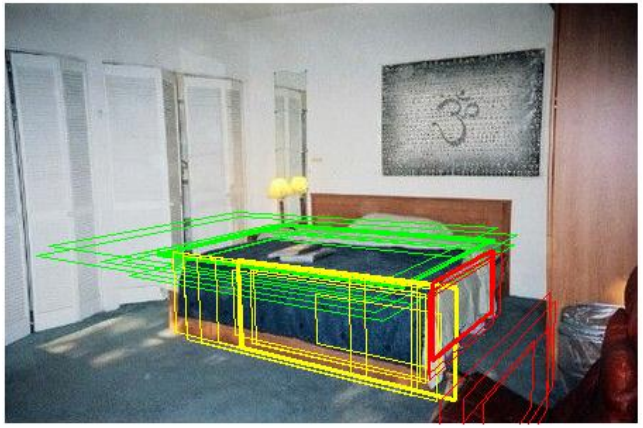
Searching for beds in room coordinates



Recover Room Coordinates

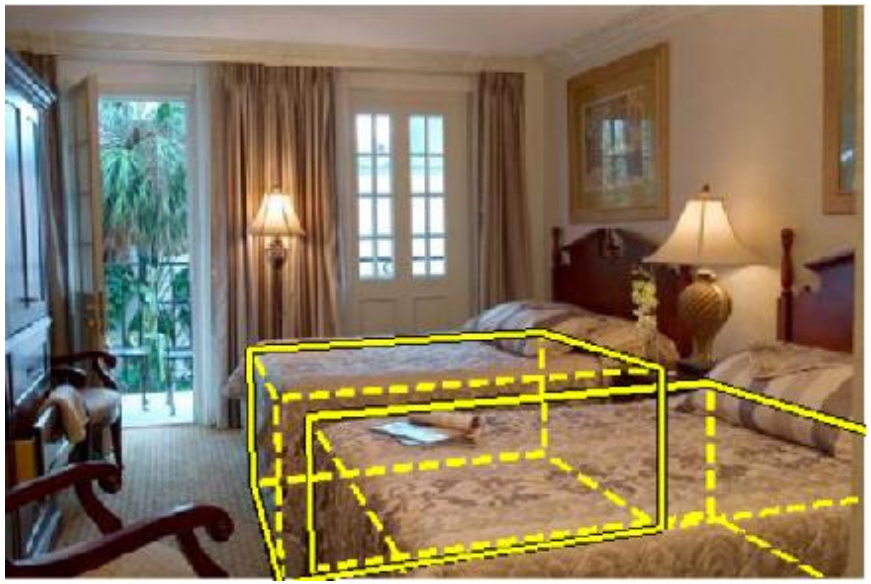


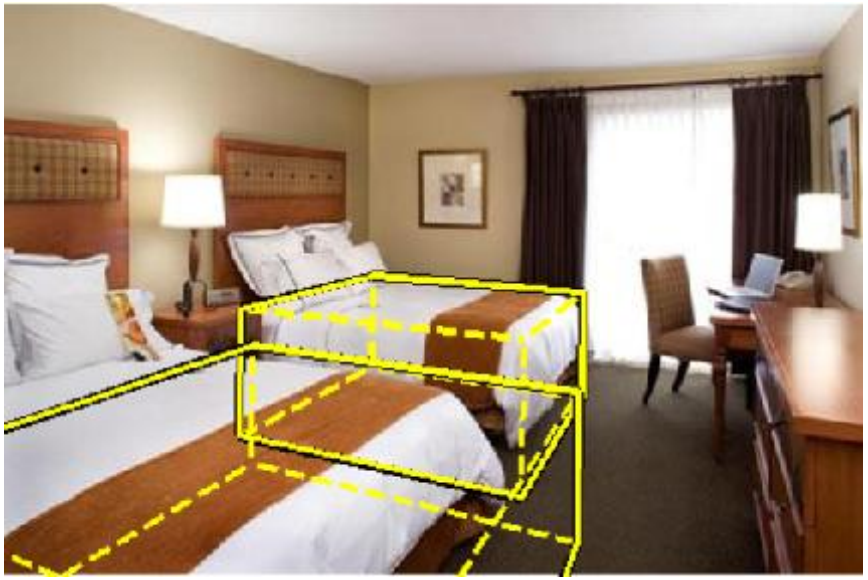
Rectify Features to Room Coordinates



Rectified Sliding Windows

3D bed detection from an image

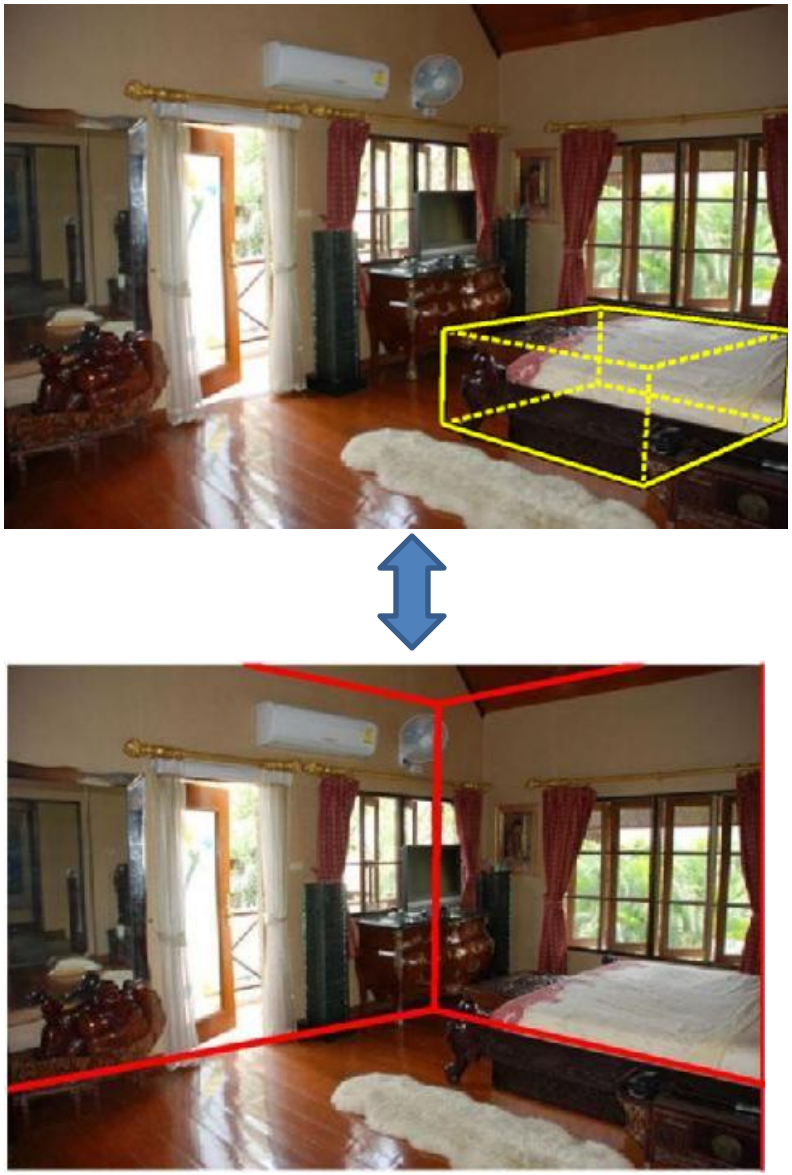




Reason about 3D room and bed space

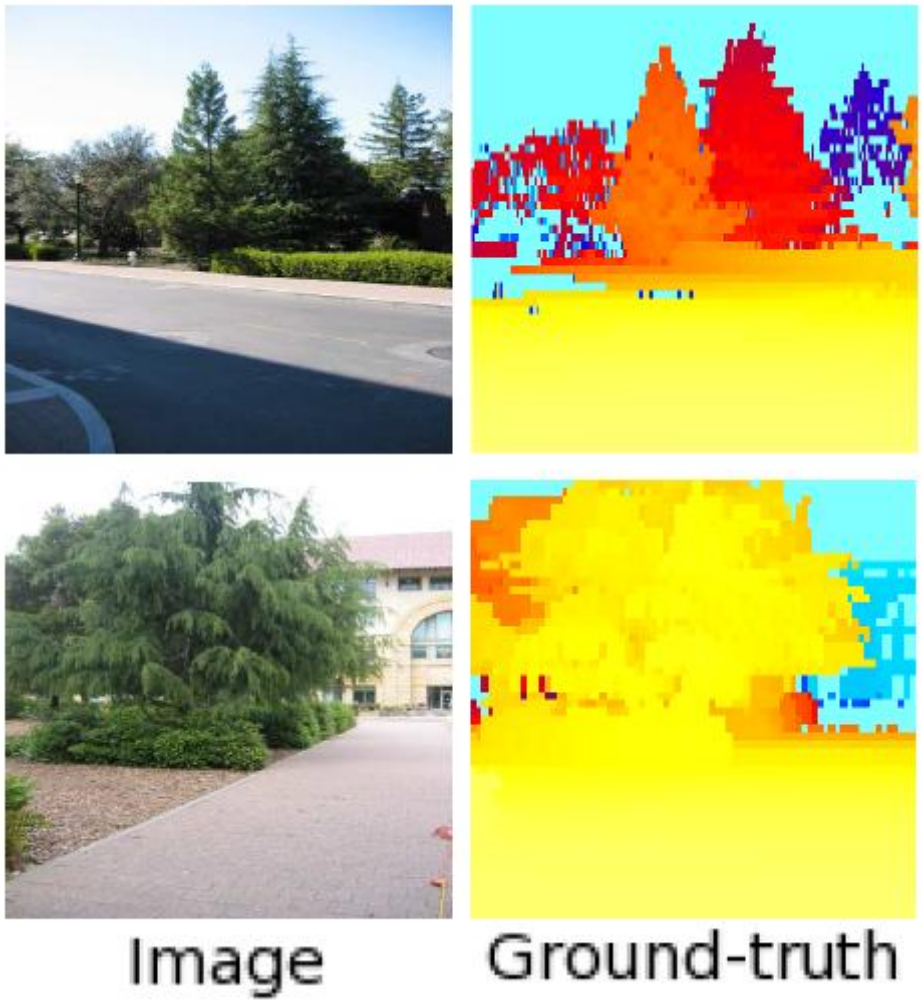
Joint Inference with Priors

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



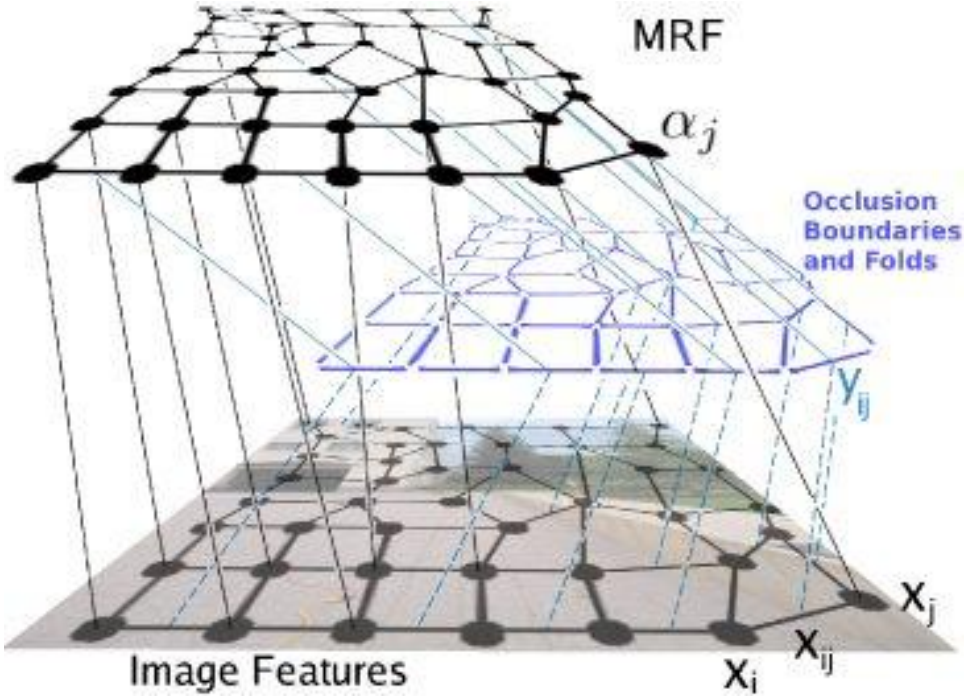
Hedau Hoiem Forsyth (2010)

Depth Estimates from an Image



Saxena et al. 2005, 2008

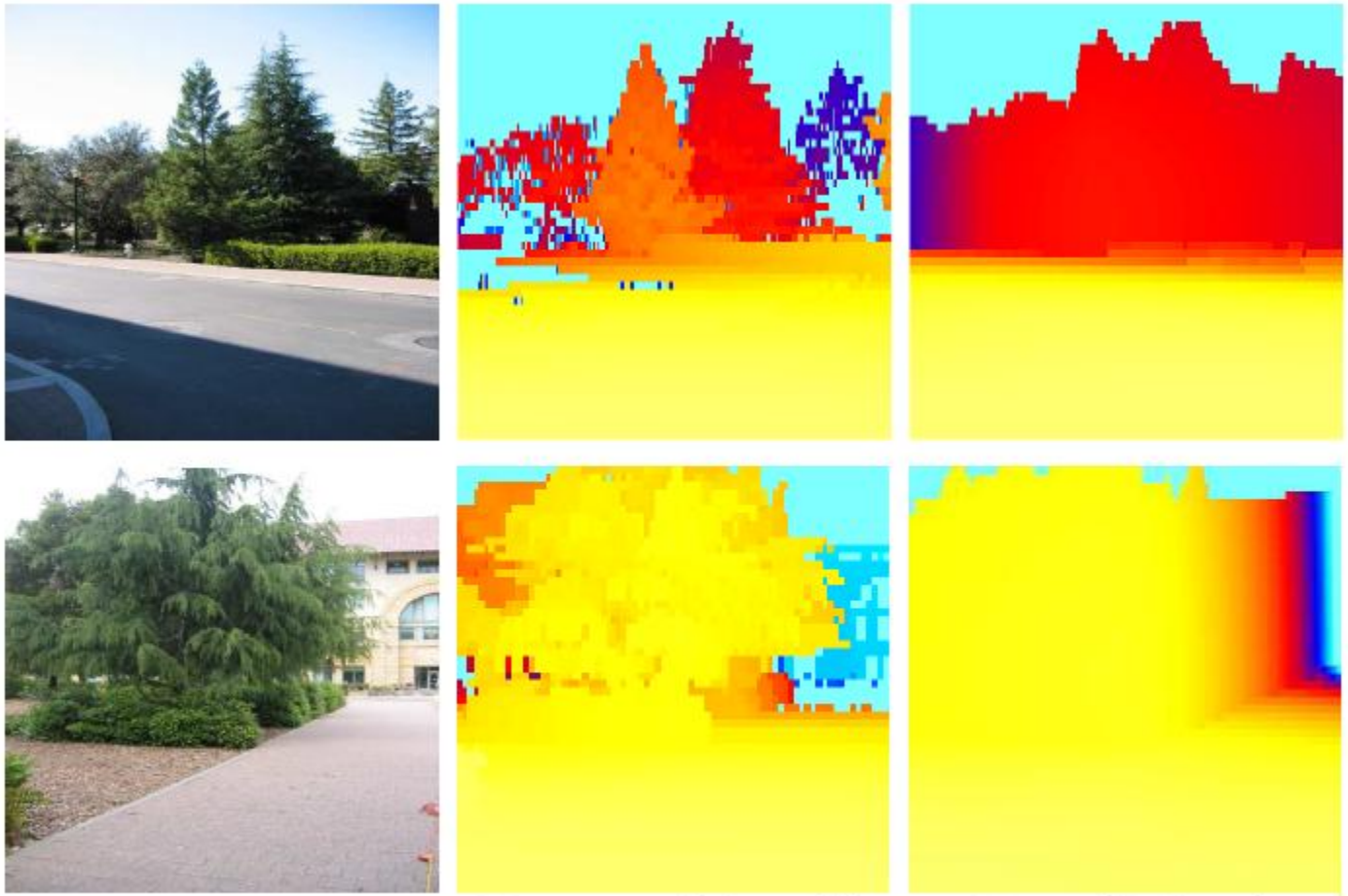
Depth from Image: approach



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

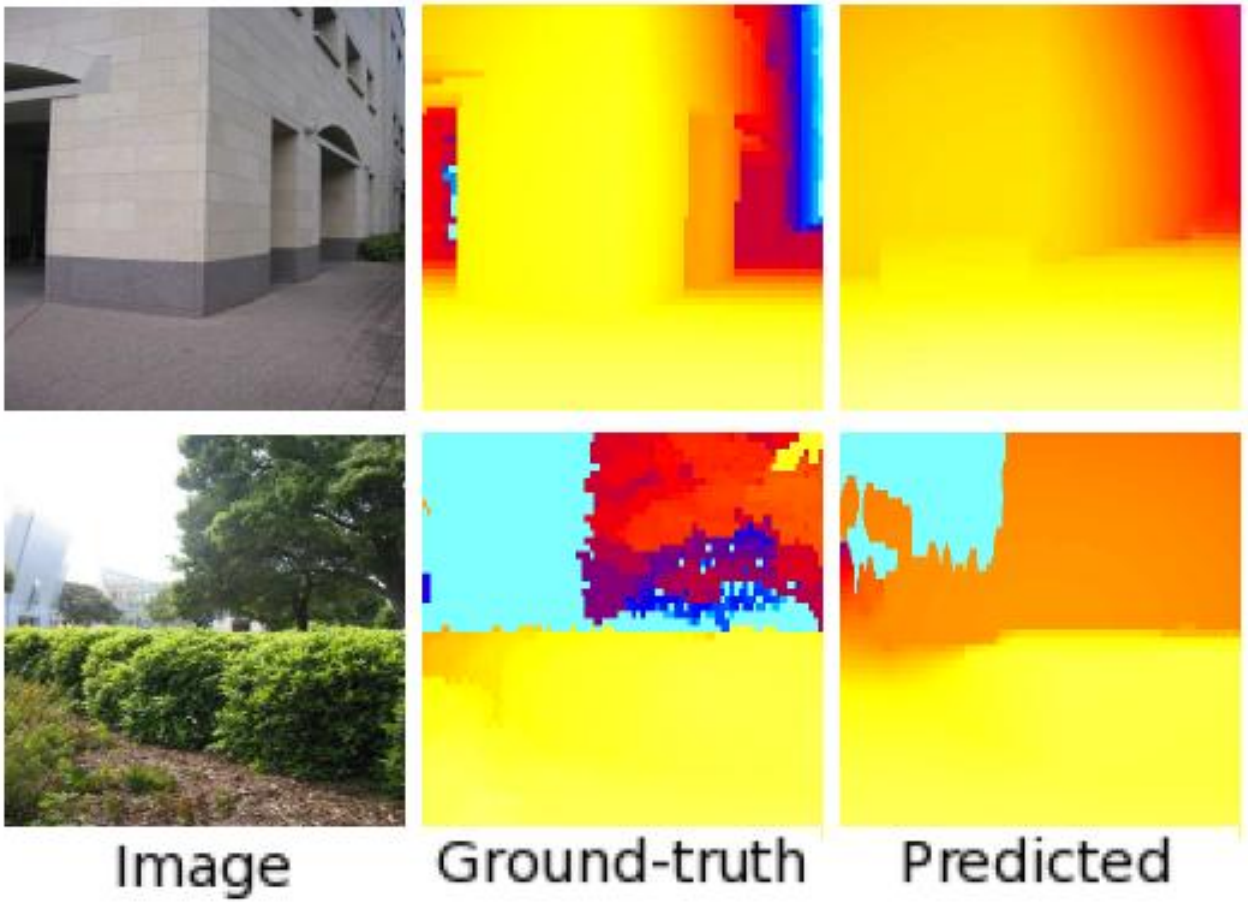
Saxena et al. 2008

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

Depth from Image

- Demo and publications:

  http://make3d.cs.cornell.edu/

- Autonomous driving (Michels Saxena Ng 2005)

Slide: Derek Hoiem

What if we had trustworthy (although coarse) geometry information?

- Rendering Synthetic Objects into Legacy Photographs. Kevin Karsch, Varsha Hedau, David Forsyth, Derek Hoiem. SIGGRAPH Asia 2011

- Project page: http://kevinkarsch.com/publications/sa11.html

- Video

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