

Context and Spatial Layout

Computer Vision
CS 543 / ECE 549
University of Illinois

Derek Hoiem

Final projects

- See class web page for details
- Poster session: May 10, 10:30-1pm
 - Will send out general invite
 - Each person presents for ~1 hour, then look at other half of posters and review two
- Paper: 4-8 pages (NIPS format)
 - Turn in at poster session or at my office by May 11 noon
 - Cannot accept late papers (or posters)

Today's class: Context and 3D Scenes

- Continuing advanced topics theme...
 - Goal is to give a broad overview of current research topics

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?



Context provides clues for function

- What is this?

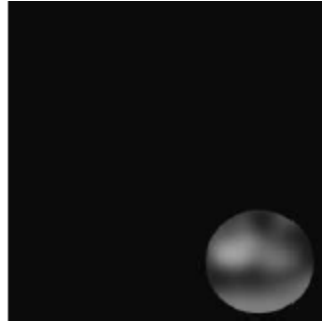


- Now can you tell?



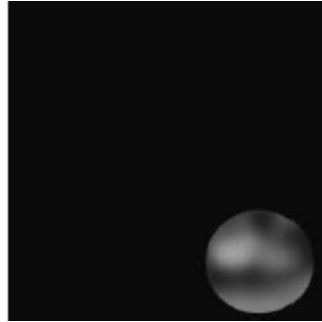
Sometimes context is *the* major component of recognition

- What is this?



Sometimes context is *the* major component of recognition

- What is this?

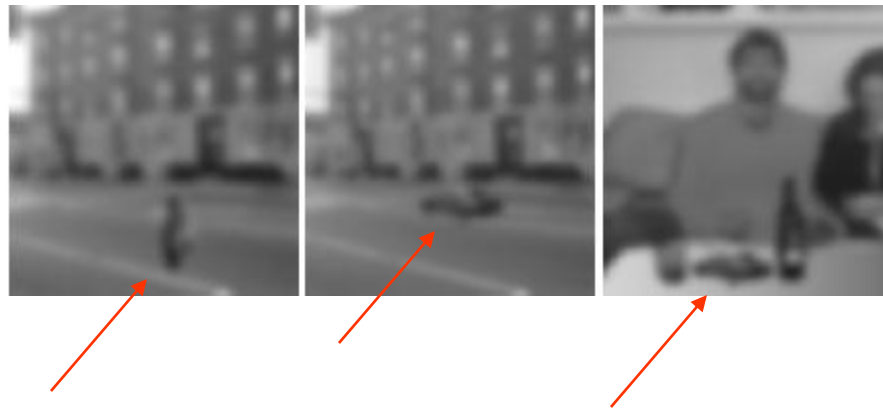


- Now can you tell?



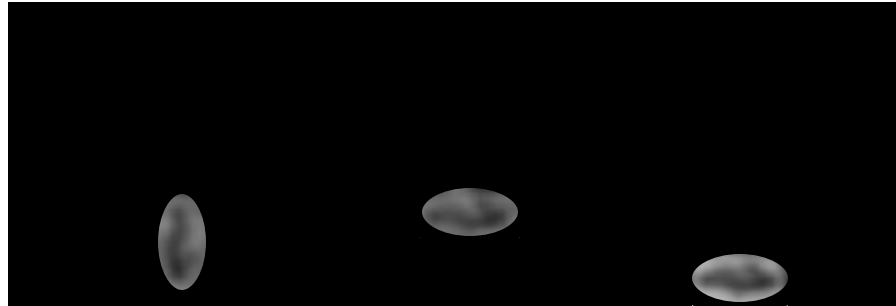
More Low-Res

- What are these blobs?



More Low-Res

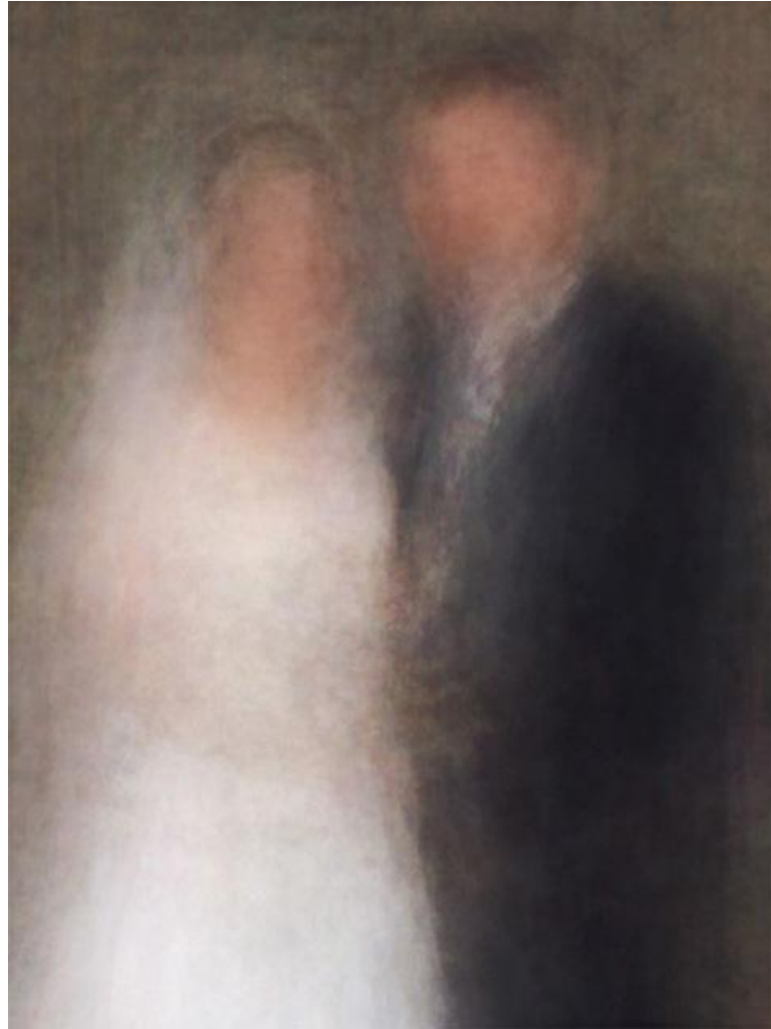
- The same pixels! (a car)



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.

Cultural context



Side note: passes Turing Test for abstract painting?



Jason Salavon: http://salavon.com/Golem/Golem_Table.shtml

Cultural context



Who is Mildred? Who is Lisa?

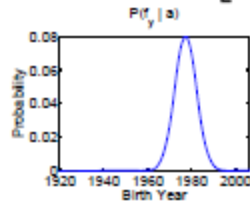
Cultural context

Age given Appearance

Age given Name

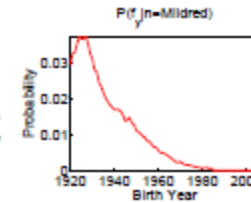


$$P(f_g|f_a) = \begin{bmatrix} 0.563 \\ 0.437 \end{bmatrix}$$



Mildred

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



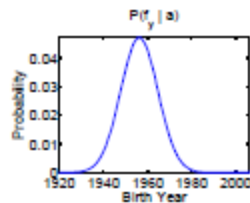
3.88

3.88

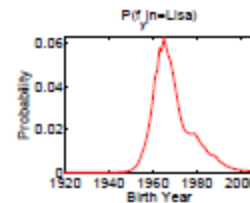
4.77

Lisa

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



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



6.70

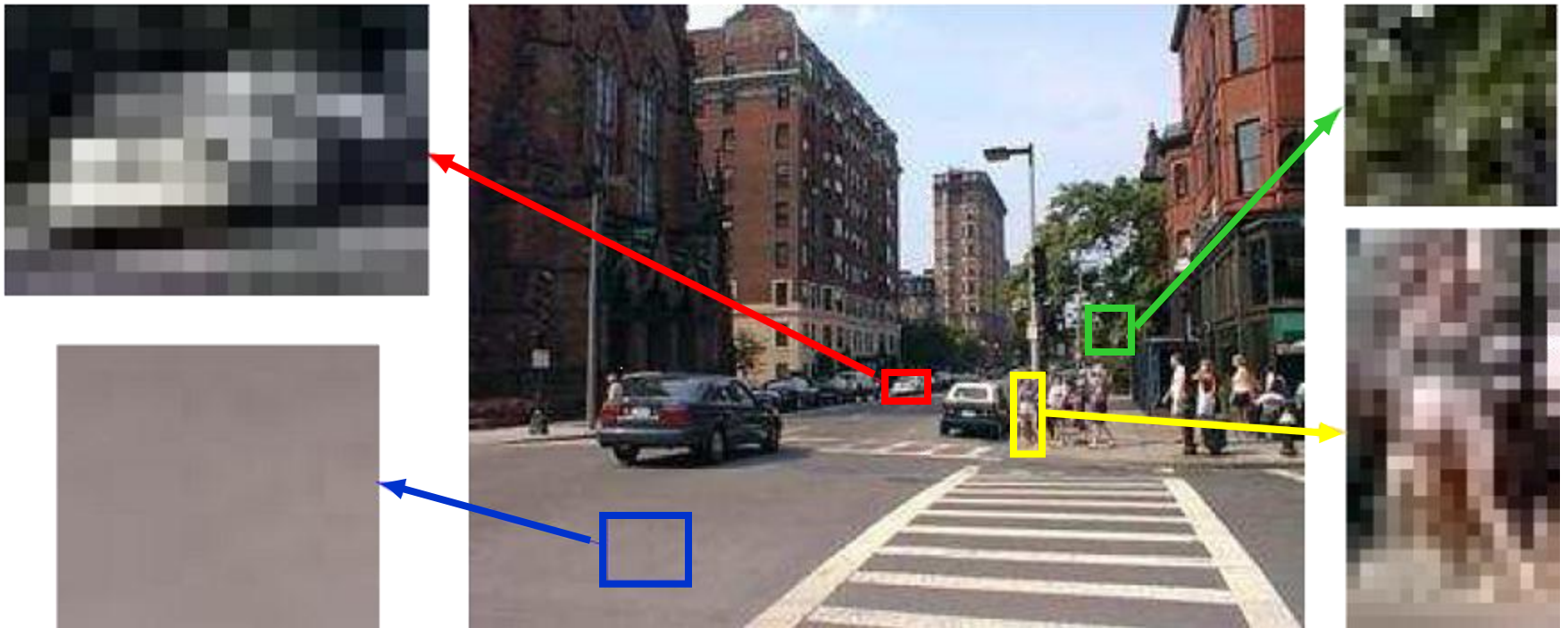
Spatial layout is especially important

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

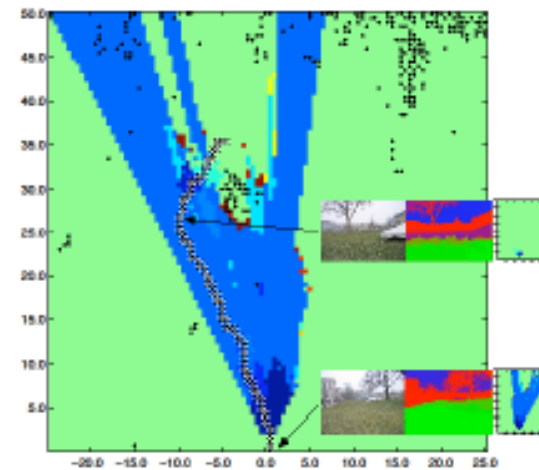


Spatial layout is especially important

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

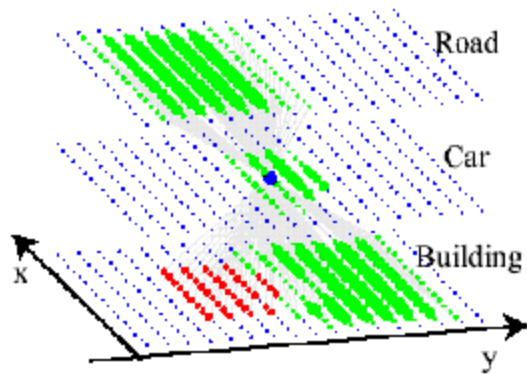


Robot Navigation: Path Planning

Spatial Layout: 2D vs. 3D

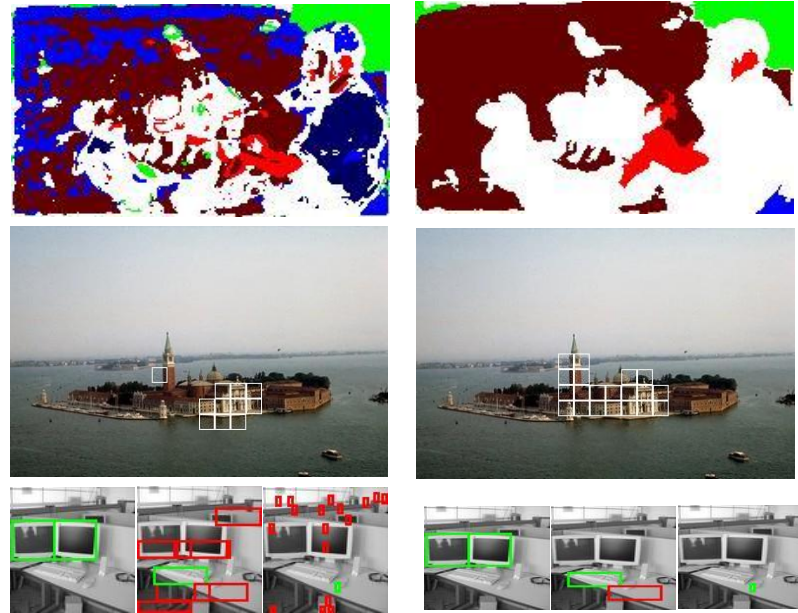


Context in Image Space

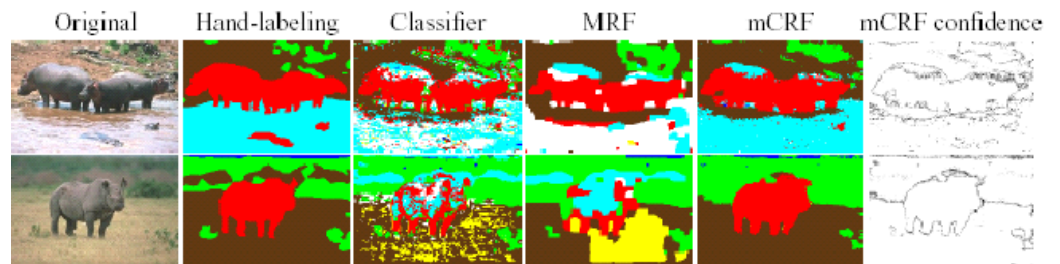


[Torralba Murphy Freeman 2004]

21

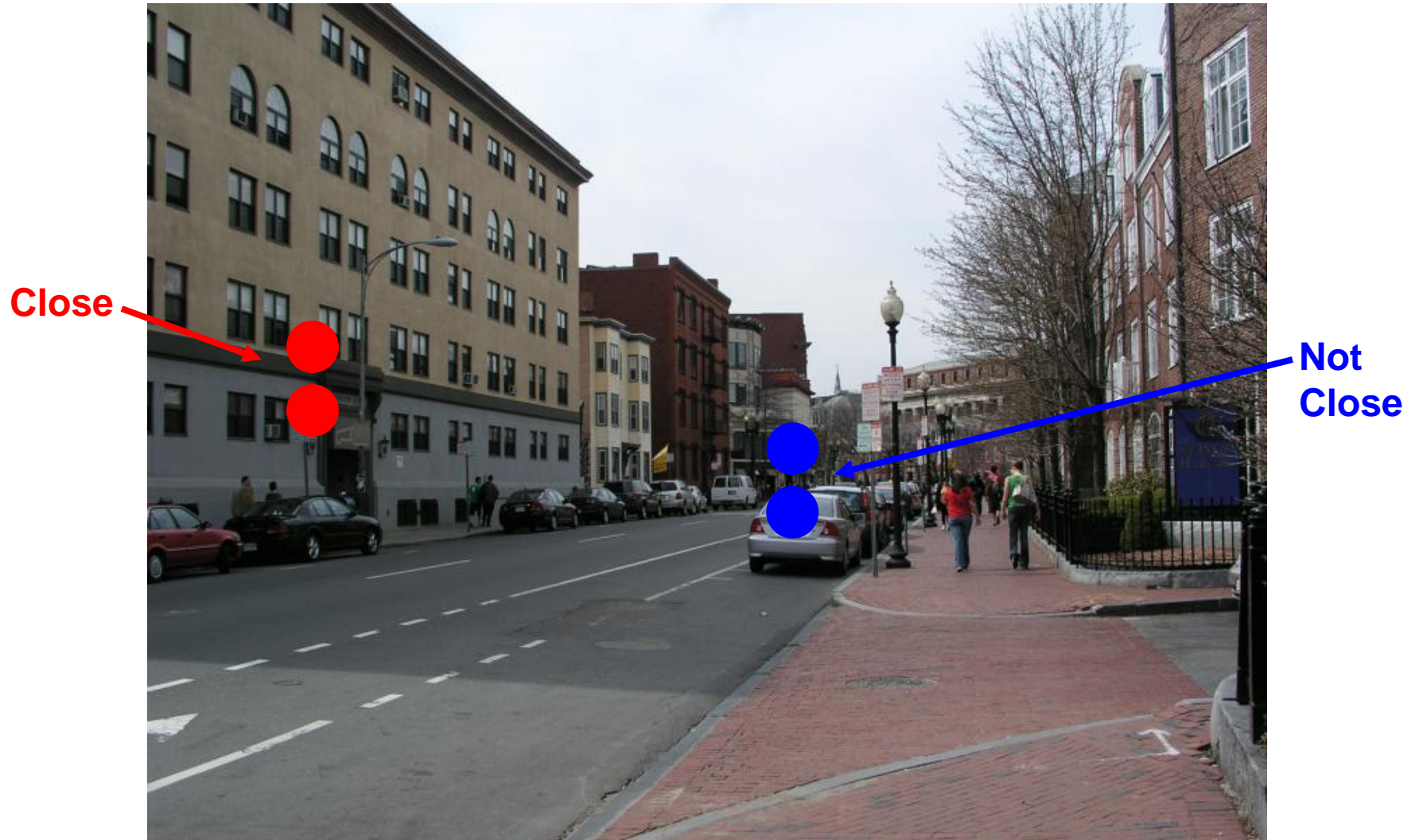


[Kumar Hebert 2005]



[He Zemel Cerreira-Perpiñán 2004]

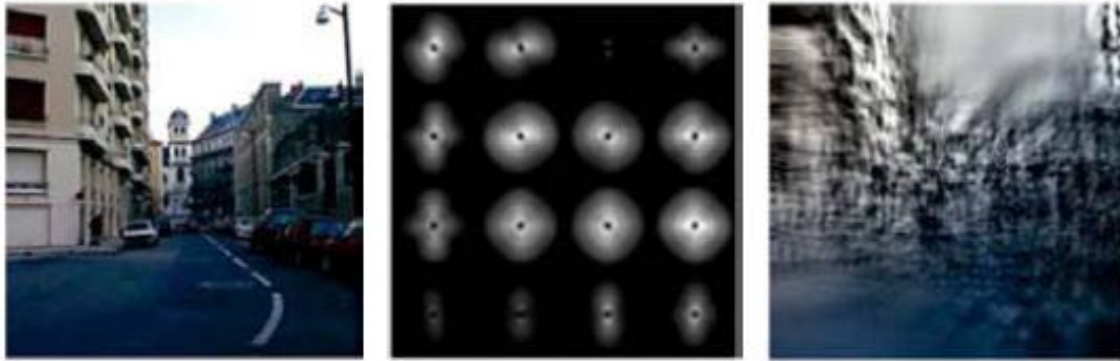
But object relations are in 3D...



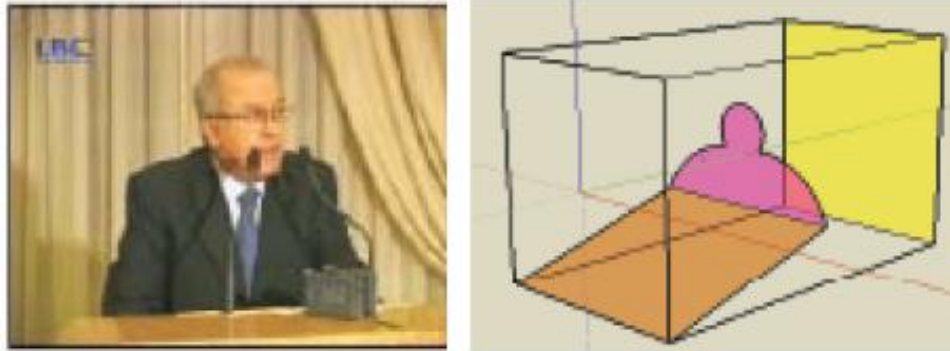
How to represent scene space?

Wide variety of possible representations

Scene-Level Geometric Description

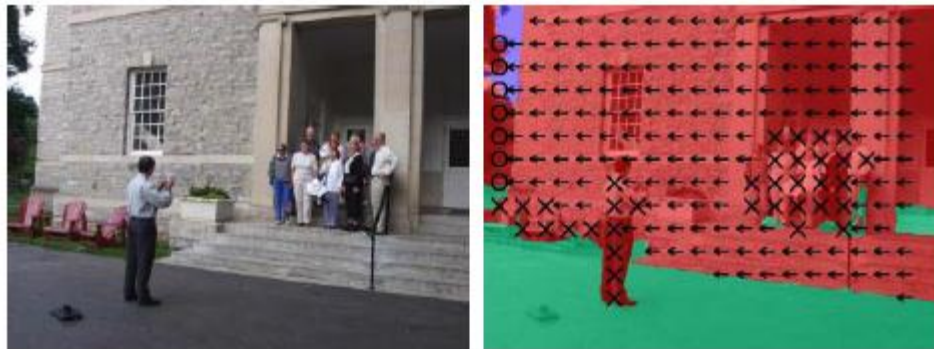


a) Gist, Spatial Envelope



b) Stages

Retinotopic Maps



c) Geometric Context



d) Depth Maps

Highly Structured 3D Models



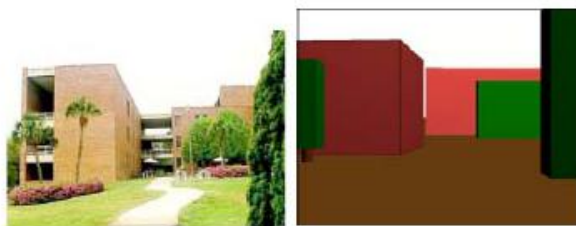
e) Ground Plane



f) Ground Plane with Billboards



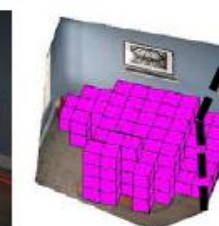
g) Ground Plane with Walls



h) Blocks World



i) 3D Box Model

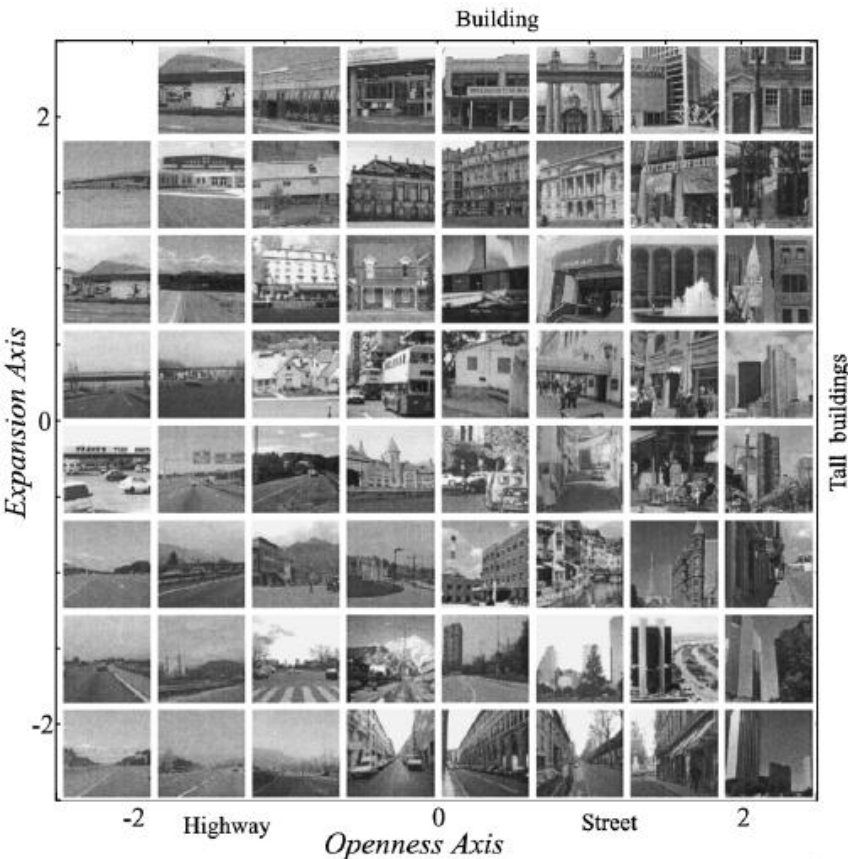


Key Trade-offs

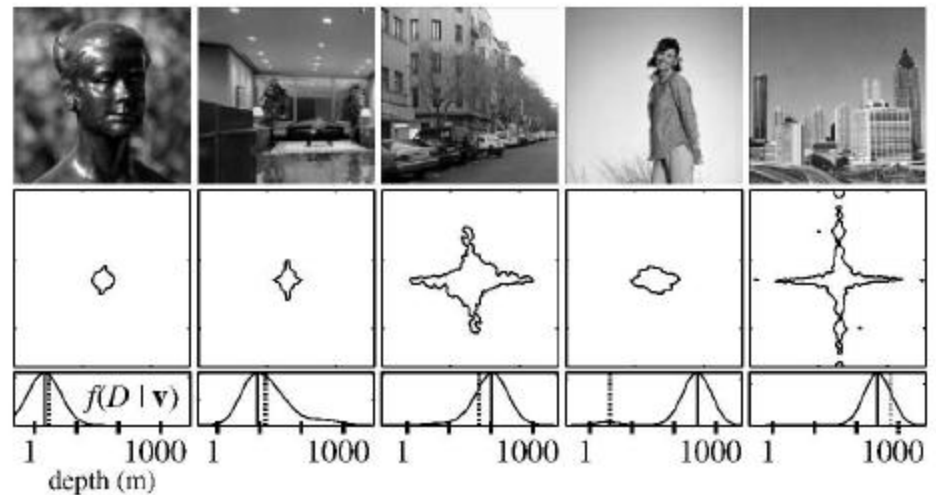
- Level of detail: rough “gist”, or detailed point cloud?
 - Precision vs. accuracy
 - Difficulty of inference
- Abstraction: depth at each pixel, or ground planes and walls?
 - What is it for: e.g., metric reconstruction vs. navigation

Low detail, Low abstraction

Holistic Scene Space: “Gist”



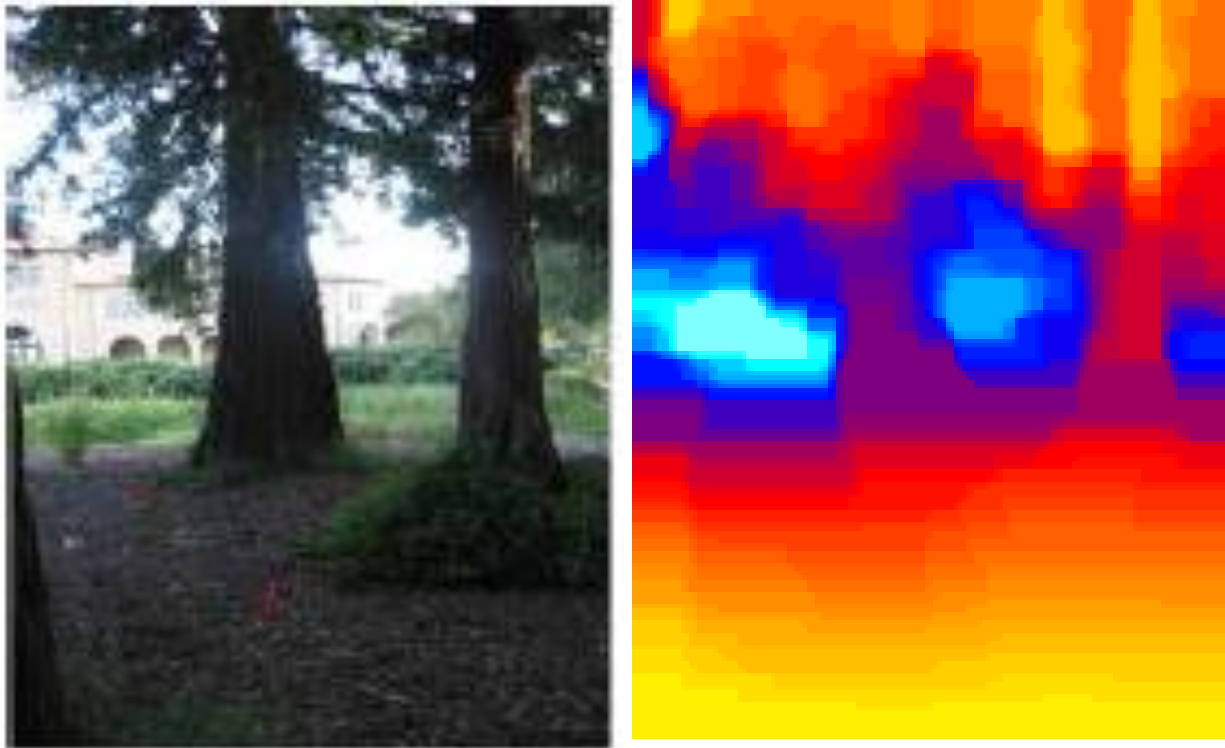
Oliva & Torralba 2001



Torralba & Oliva 2002

High detail, Low abstraction

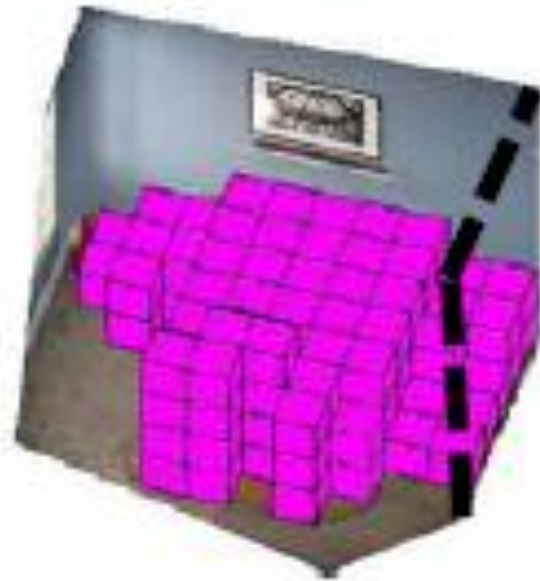
Depth Map



Saxena, Chung & Ng 2005, 2007

Medium detail, High abstraction

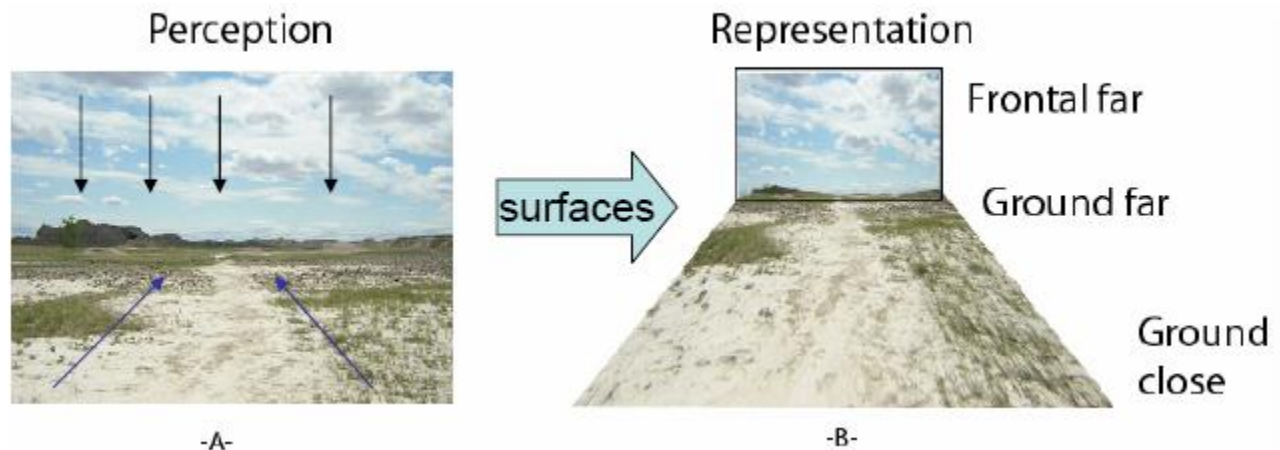
Room as a Box



Hedau Hoiem Forsyth 2009

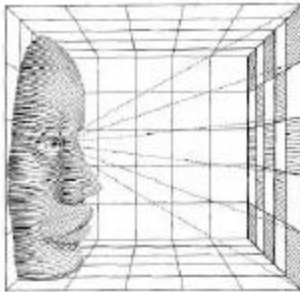
Gibson's Surface Layout

- 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

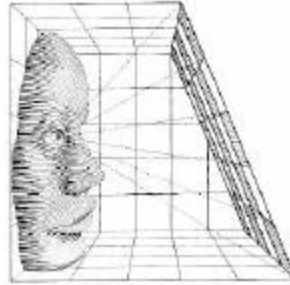


Gibson's Surface Layout

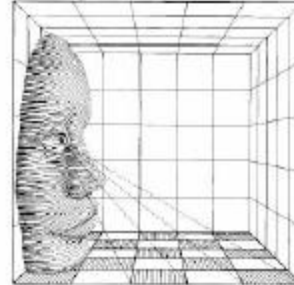
Flat frontal vertical surface



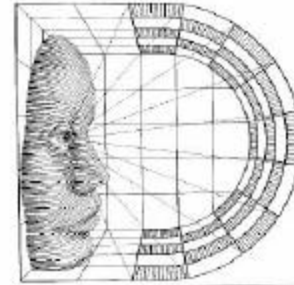
Flat frontal slanting surface



Flat longitudinal ground surface



Rounded surface



Gibson's Surface Layout



Vertical surface texture

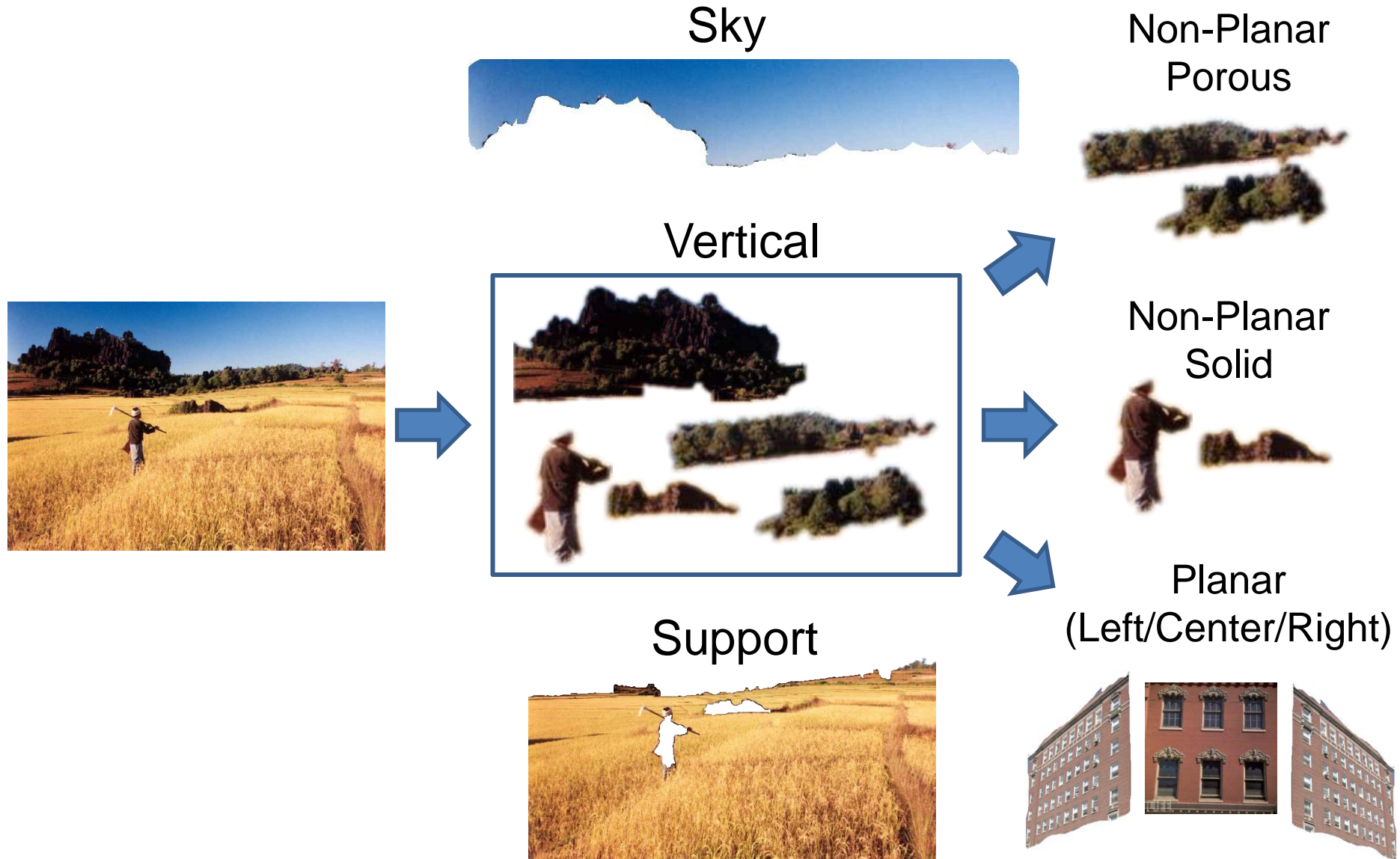


Ground surface texture

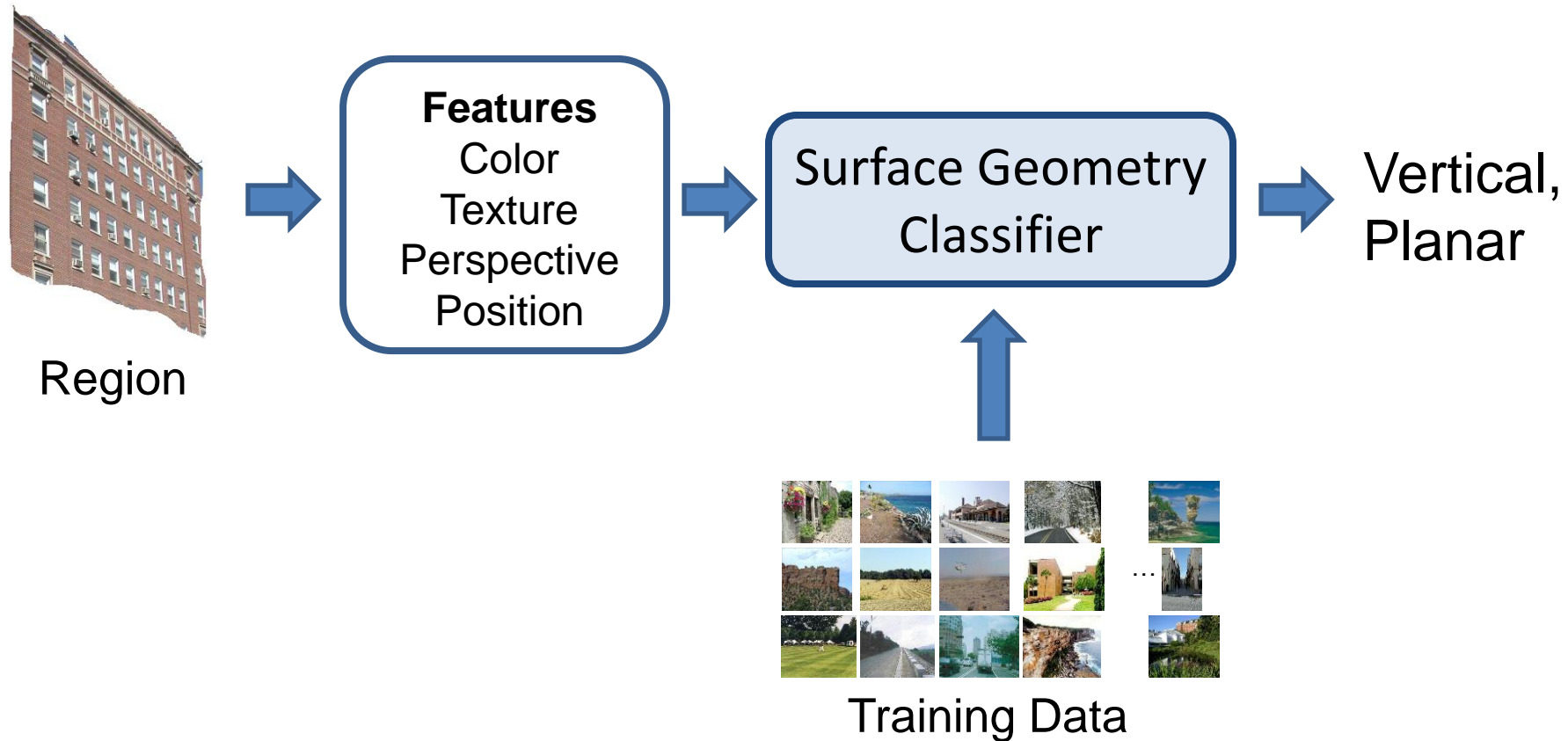
A few examples of spatial layout estimation

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

Surface Layout: describe 3D surfaces with geometric classes



Geometry estimation as recognition



Use a variety of image cues



Vanishing points, lines

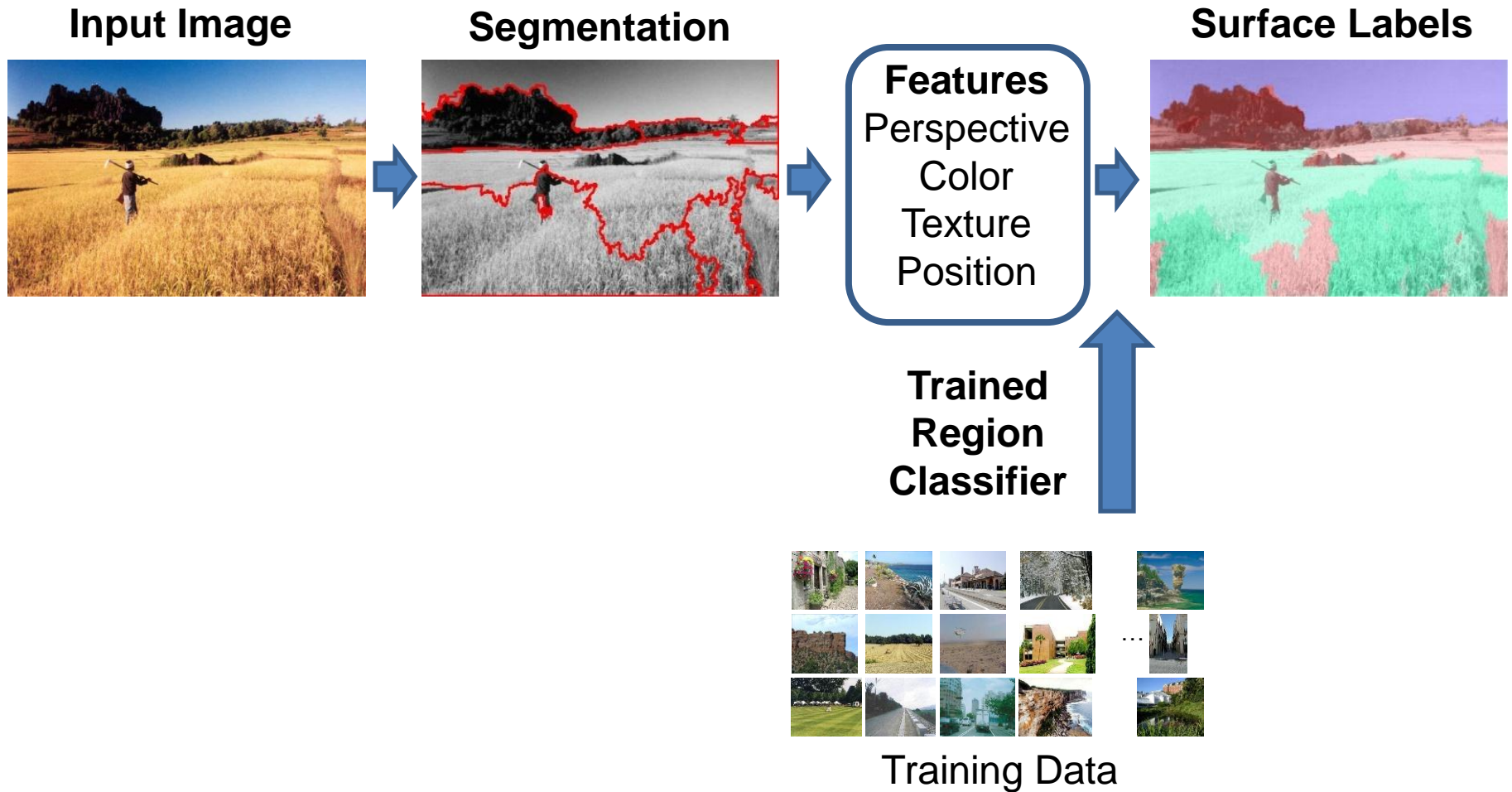


Color, texture, image location

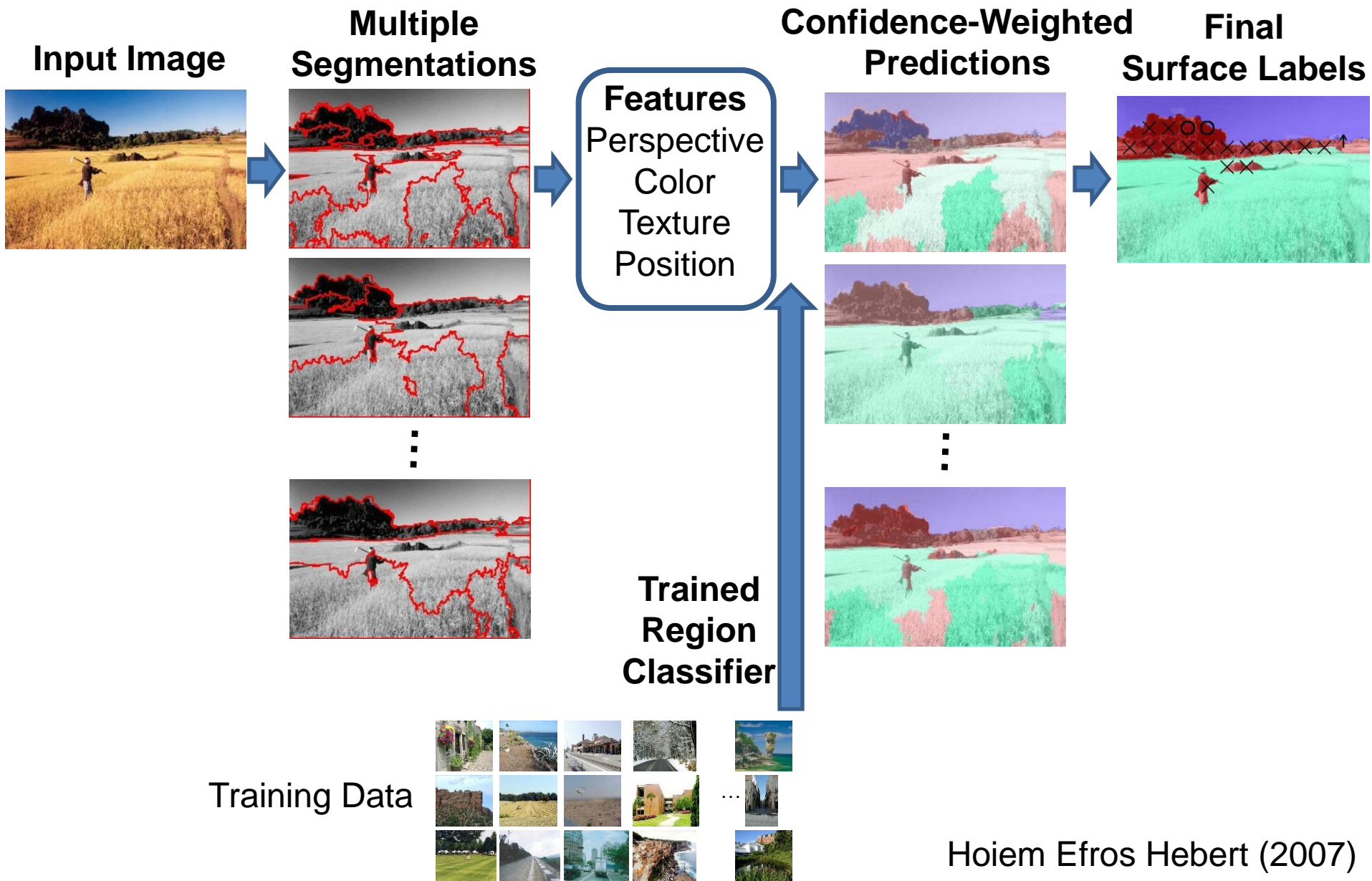


Texture gradient

Surface Layout Algorithm



Surface Layout Algorithm



Surface Description Result



Automatic Photo Popup

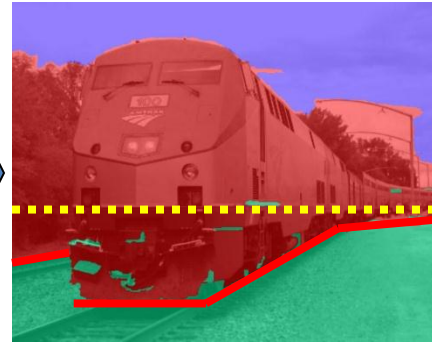
Labeled Image



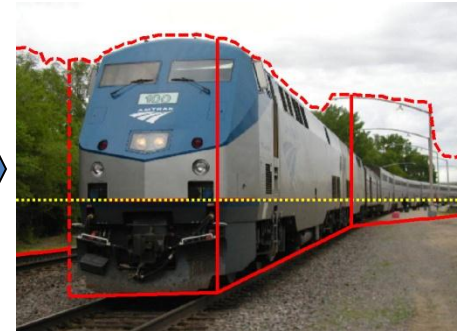
Fit Ground-Vertical
Boundary with Line
Segments



Form Segments
into Polylines



Cut and Fold



Final Pop-up Model



Automatic Photo Pop-up





What about more organized but complex spaces?



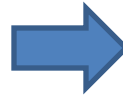
Other excellent works include:

Saxena Sun Ng (2009)

Lee Kanade Hebert (2009)

Gupta Efros Hebert (2010)

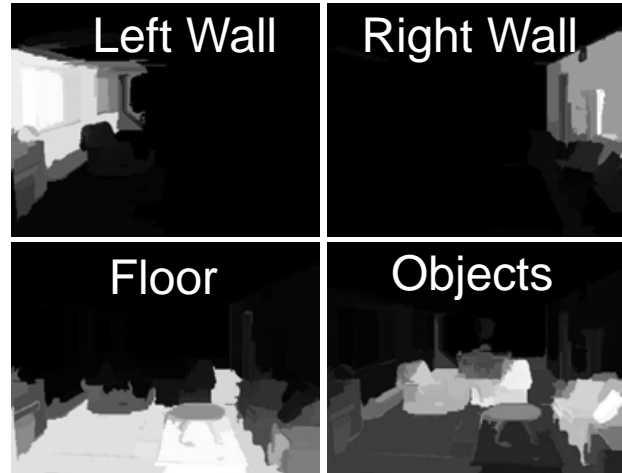
The room as a box



Recovering the box layout

Vertical
VP 

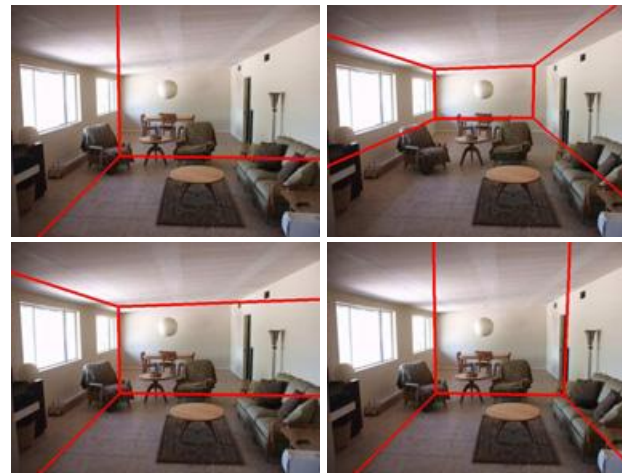
Surface Label Confidences



Detected Edges
+ Vanishing Points



Joint
Inference

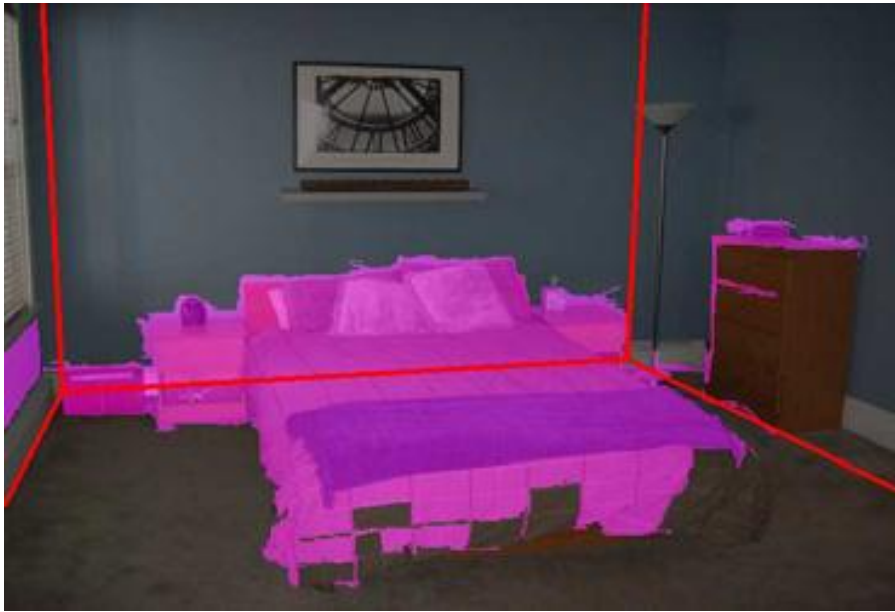


Hypothesized Boxes

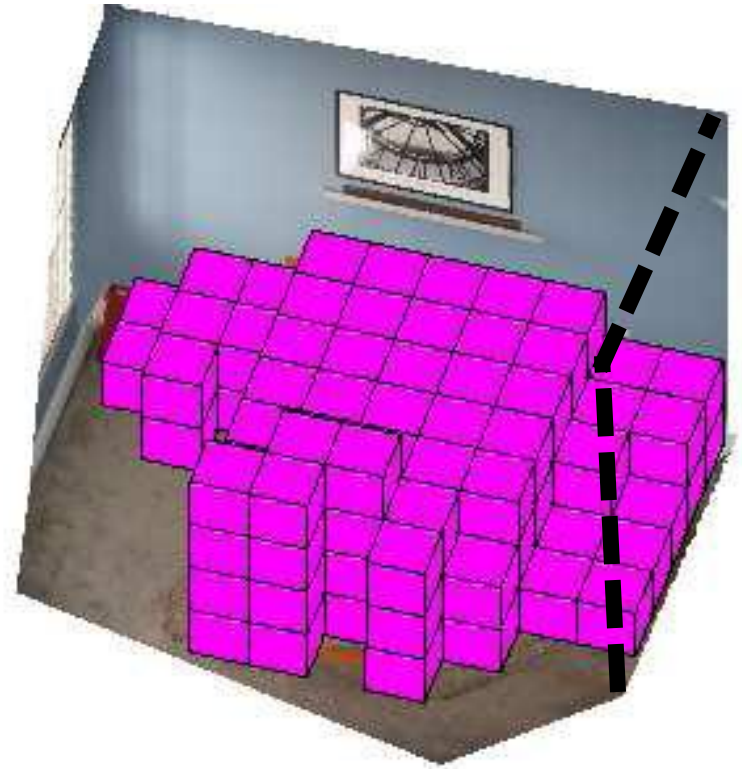


Most Likely
Geometry

Estimate room's physical space from one image



Estimated “Box” Geometry
+ Object Pixels



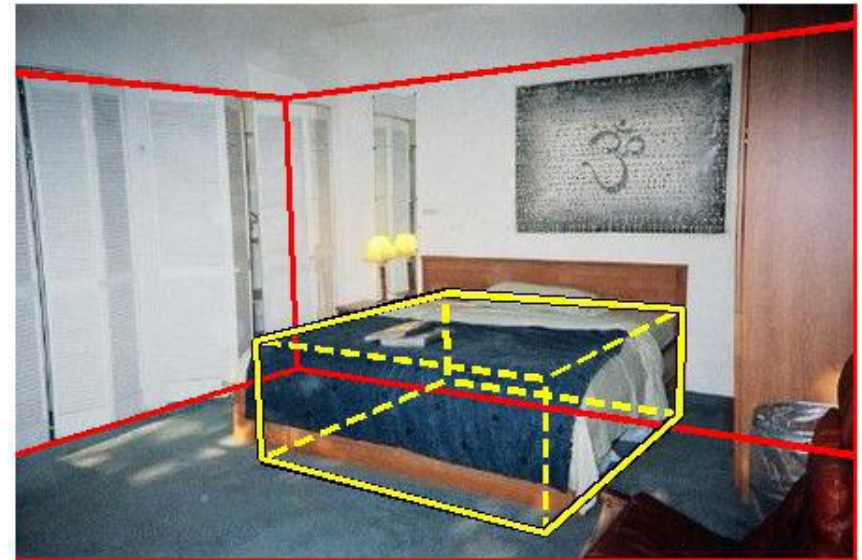
3D Reconstruction +
Estimated Occupied Volume

Detecting 3D bed positions in an image

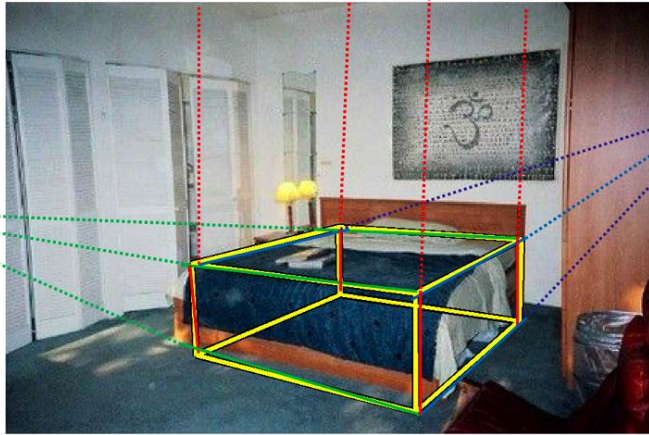
2D Bed Detection



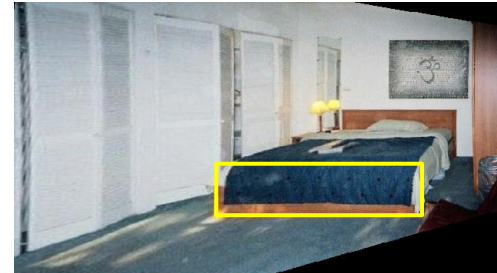
3D Bed Detection with Scene Geometry



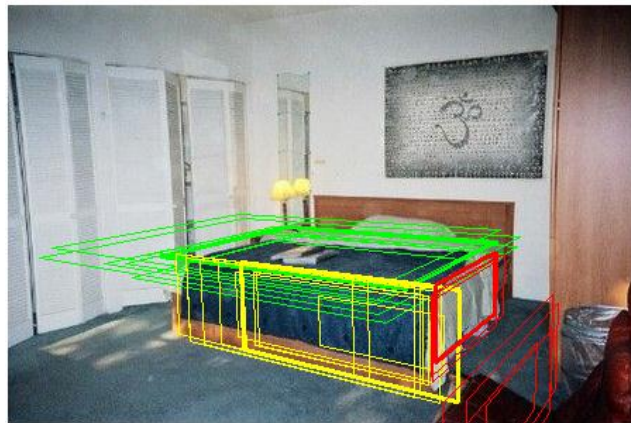
Searching for beds in room coordinates



Recover Room
Coordinates

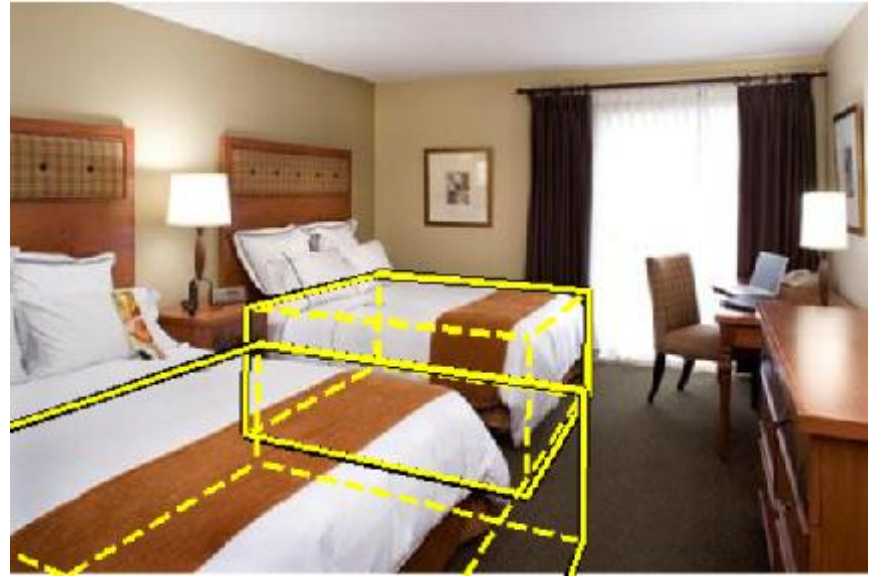
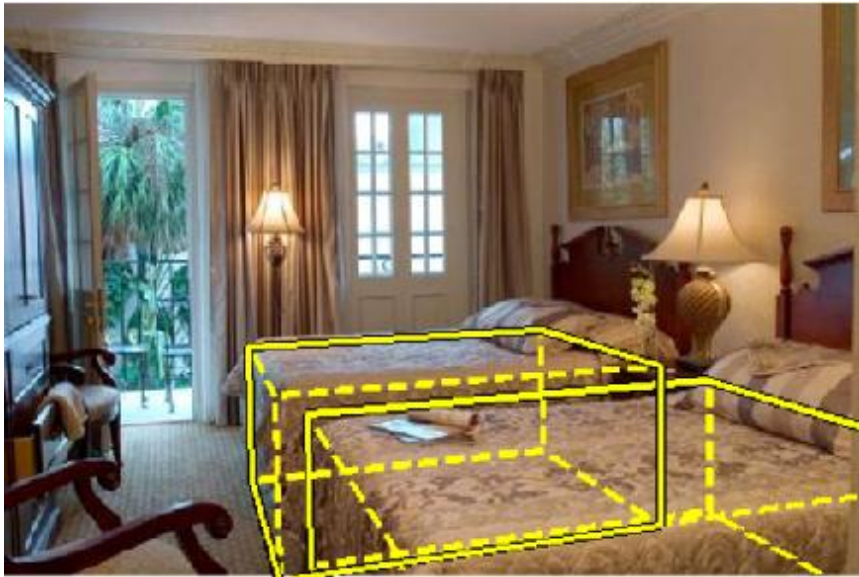


Rectify Features to Room
Coordinates



Rectified Sliding Windows

3D bed detection from an image



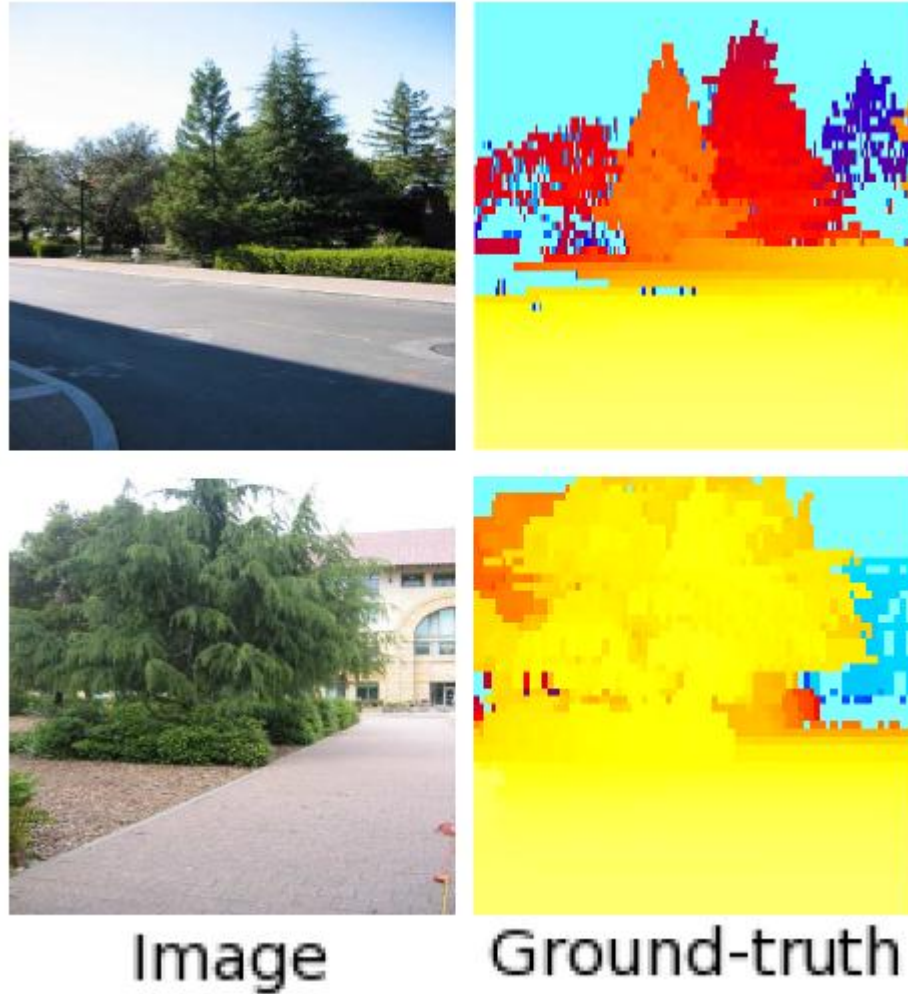
Reason about 3D room and bed space

Joint Inference with Priors

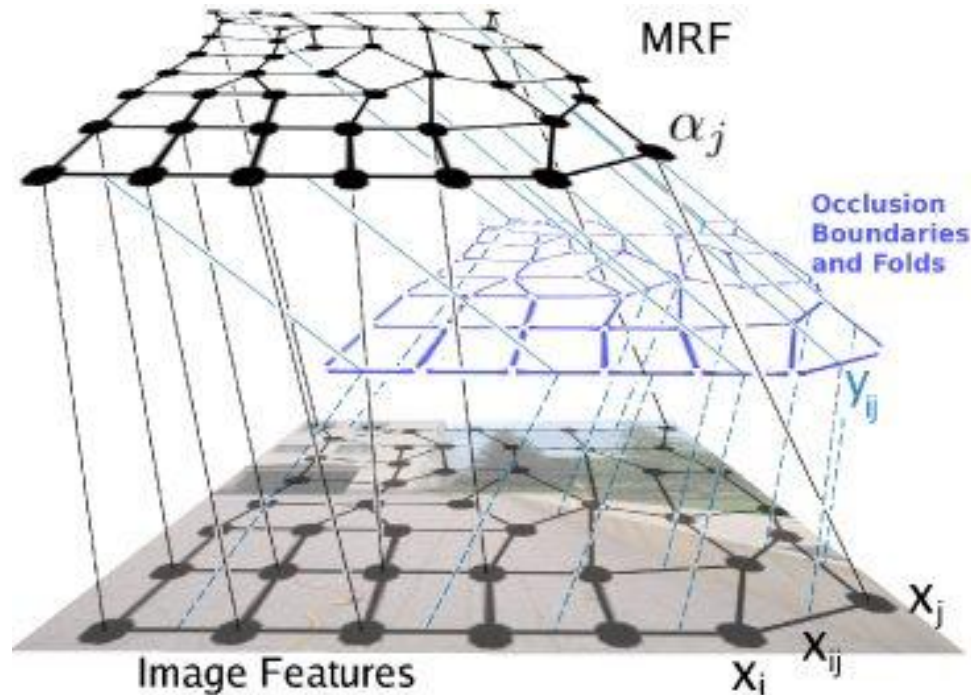
- Beds close to walls
- Beds within room
- Consistent bed/wall size
- Two objects cannot occupy the same space



Depth Estimates from an Image

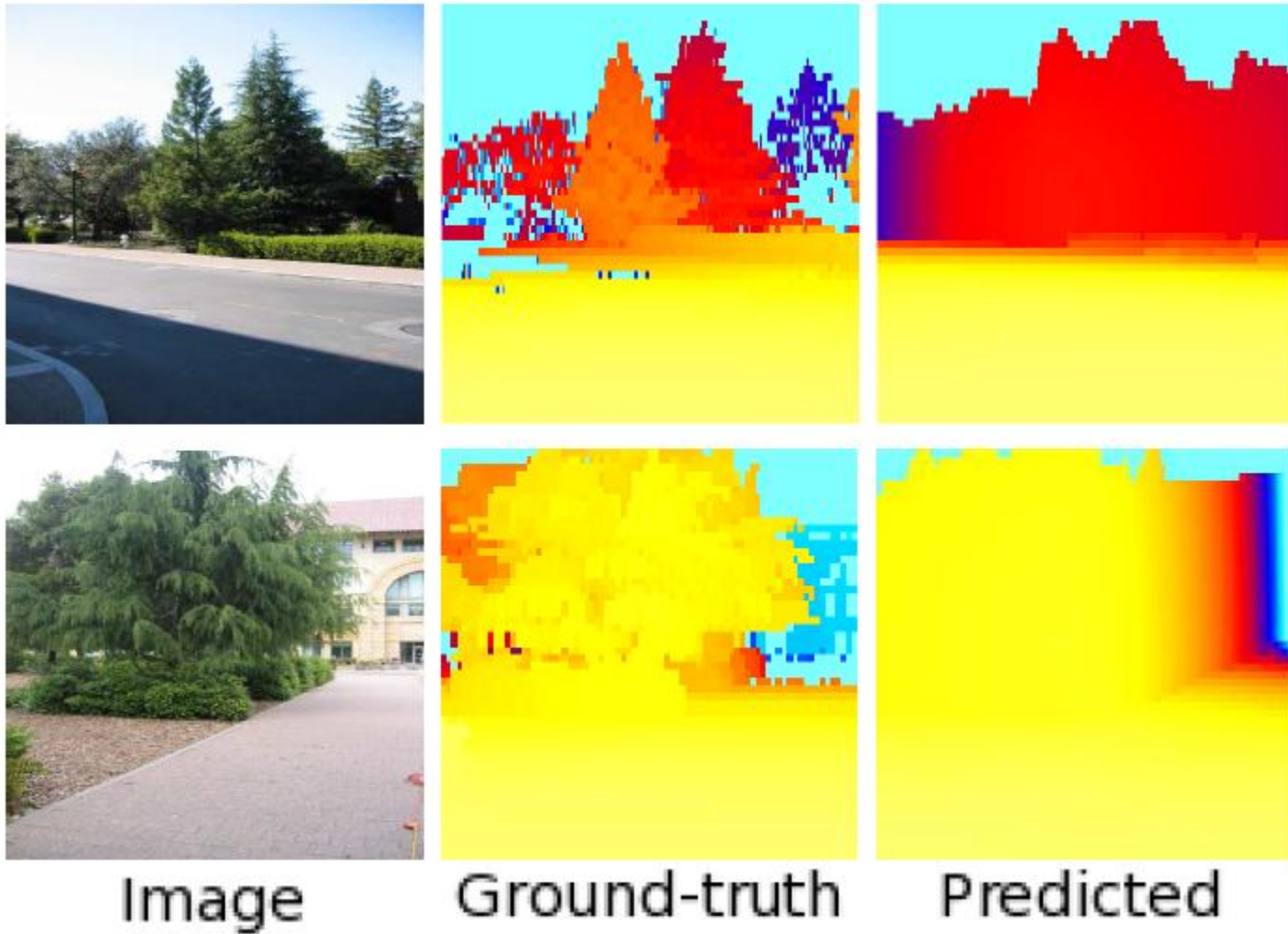


Depth from Image: approach

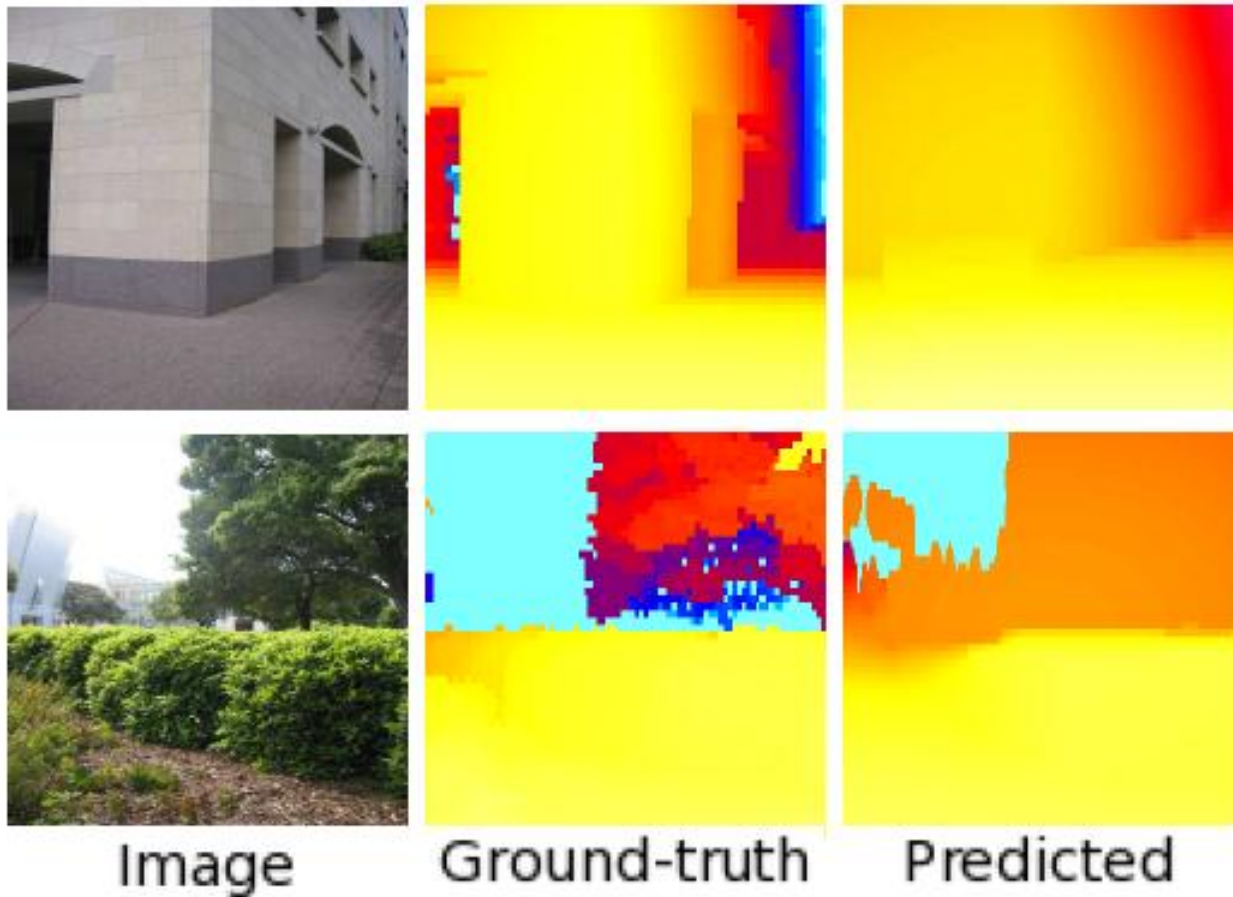


1. Divide image into superpixels
2. Compute features for each superpixel
 - Position, color, texture, shape
3. Predict 3D plane parameters for each superpixel using features
4. Estimate confidence in prediction using features
5. Global inference, incorporating constraints of connected structure, co-planarity,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

Depth Estimates from an Image



Depth Estimates from an Image



Depth from Image: Reconstructions

Input



Novel View



Depth from Image

- Demo and publications:

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

- Autonomous driving (Michels Saxena Ng 2005)



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

Next class: what you can do with lots of images (massively data-driven approaches)



The Big Picture, Birmingham, England