More Constraints from Objects and Actions

Abhinav Gupta
Carnegie Mellon University
Recap
Goal: Single Image 3D
Kanade’s chair... (Artificial Intelligence, 1981)
Applications
Object Detection
Autonomous Driving
Qualitative

Region labels
3D Scene Understanding

Hoiem et al. (2005)
Saxena et al. (2007)
Region labels

Constraints from domain knowledge

Qualitative

More Reasoning
Can we include extra constraints in the scene understanding process?
BUT..
From Planes to Volumes
Benefits of Volumes

Geometrical Relationships

Roberts (1965), Guzman (1968)

Mechanical Relationships (Physical Stability)

Blum (1970), Winston (1972)
Left-Occluded

Supported by

Left-Right

infront-of

Medium-Heavy

Point-supported

infront-of

above

Heavy

3D Parse Graph

Blocks World Revisited: Image Understanding using Qualitative Geometry and Mechanics. A. Gupta, A.A. Efros, M. Hebert. ECCV 2010
Qualitative Volumes

Catalogue

15 m

10 m
Qualitative Volumes

Catalogue
Qualitative Densities

Light

Heavy
• Three **Qualitative Classes**: Light, Medium and Heavy
• Appearance and Location Features.
• Decision Tree Classifier

<table>
<thead>
<tr>
<th>Appearance</th>
<th>Location &amp; Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color (RGB)</td>
<td>Mean (x, y)</td>
</tr>
<tr>
<td>Color (HSV)</td>
<td>x, y (10\textsuperscript{th} percentile)</td>
</tr>
<tr>
<td>Hue, Saturation Histograms</td>
<td>x, y (90\textsuperscript{th} percentile)</td>
</tr>
<tr>
<td>DOOG Filters</td>
<td>Size of Segments</td>
</tr>
<tr>
<td>DOOG Statistics</td>
<td>Convexity</td>
</tr>
</tbody>
</table>

Light

Medium

Heavy
Goal

Hard Combinatorial Optimization
Building Blocks World
Surface Layout Density Map

Round 1

Bag of Segments

Catalogue

Frontal Front-Right Front-Left

Left-Right Left-Occluded Right-Occluded

Porous Solid
Bag of Segments

Surface Layout Density Map

Round 1

Catalogue

Frontal Front-Right Front-Left

Left-Right Left-Occluded Right-Occluded

Porous Solid

Ground
Fitting Cuboids

Input Images
Building 3D Blocks World

Input Images

Toy Blocks World Rendering
## Quantitative Evaluation

<table>
<thead>
<tr>
<th>Approach</th>
<th>Surface Layout</th>
<th>Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hoiem (2008)</td>
<td>68.8%</td>
<td>0.65</td>
</tr>
<tr>
<td>Our Approach</td>
<td>73.7%</td>
<td>0.69</td>
</tr>
</tbody>
</table>
Integrating constraints in INDOOR
Constraints: Solid objects must satisfy volumetric/physical constraints

- Finite volume
- Spatial exclusion
- Containment

Method Overview

Room Hypotheses

Object Hypotheses
Object Hypotheses - Surfaces

Image

Orientation Map

Convex Edge Check

Hypothesized Cuboid

Object Hypotheses
Object Hypotheses - Corners

Understanding Bayesian rooms using composite 3D object models. L. Del Pero, J. Bowdish, E.L. Hartley, B. Kermgard, K. Barnard. CVPR 2013
Object Hypotheses – Contact Points
Method Overview

- Room Hypotheses
- Object Hypotheses
- Test Volumetric Constraints
- Geometric context [Hoiem05]
- Orientation map
- Evaluate
- Scene configuration hypotheses
- Final scene configuration
• What is the right Evaluation Function?

• How does one search the space of possible hypothesis?
Evaluation Function

- Discriminative Approach
- Generative Approach
Discriminative Approach

\[ \text{score(img, config)} = w^T f(img, config) \]
Weight learning: Structured SVM

\[ f(x, y) = w^T \psi(x, y) + w_\phi^T \phi(y) \]

- Features from image (surface labels, vanishing points, etc.)
- Features of the scene configuration to evaluate constraint violations
- Compatibility of image data with geometric configuration
- Penalty term for incompatible configurations
- Weight learning: Structured SVM [Tsochantaridis05]
Loss Function: Number of pixels labeled wrong
Generative Approach

Represent the Generative Model of generating edges, geometric context, orientation map...given the room and object hypothesis.

Hypothesis \( \theta \) = *Both Scene + Camera*

\[
\theta = (r, c)
\]

Scene model: \( r = (r_b, n, o_1, \ldots, o_n) \)

Camera model: \( c = (\psi, \varphi, f) \)

Goal: \( p(\theta|D) \propto p(D|\theta)p(\theta) \)
Given a 3D hypothesis

3 Possible Situation
• Edge detected at expected model location [Reward].
• Edge detected where no edge was expected [Penalty].
• Edge expected but no edge detected [Penalty].

Similar likelihood model for geometric context and other features
• What is the right Evaluation Function?

• How does one search the space of possible hypotheses?
Greedy

Top Room Hypothesis

Add Object
In a greedy fashion

Stop when score decreases
Greedy (Beam Search)

Top N Room Models
MCMC Style Sampling

Propose layout
Propose frame
block

Moves:
• Jump Moves: Add and Delete Block/Frame
• Diffusion Moves: Search over parameters of block, room, camera

Understanding Bayesian rooms using composite 3D object models. L. Del Pero, J. Bowdish, E.L. Hartley, B. Kermgard, K. Barnard. CVPR 2013
Full fit

Focal length estimation

Blocks explain occlusions
Optimal Search: Box in a Box

Complexity: Occluded Areas Should not be Counted Twice

Assumption: Only One Object

Branch and Bound Approach

Qualitative

More quantitative
more precise

Region labels
Stronger geometric constraints from domain knowledge
Physical Reasoning
Constraining Hypothesis Space

Object Hypothesis: Could be cuboid of any size..
Using relative placement statistics

Objects: statistics on relative size and contact for each object type $i$

$$r_{i1} = \frac{h_i}{\max(w_i, l_i)}$$

$$r_{i2} = \frac{\max(w_i, l_i)}{\min(w_i, l_i)}$$

$$r_{i3} = \frac{h}{h_i}$$

$d_i = 1$ if surface contact

[Example from Del Pero]
Use semantics as well

There are even stronger constraints
Generalization to groups of objects (geometric phrases)

Qualitative

Region labels

Stronger geometric constraints from domain knowledge

Physical Reasoning

Semantic Reasoning

More reasoning
Even more constraints: Functional

1. Structural model function-geometry-appearanc

2. Estimate distributions from training data

3. Sample using model

Functional and Domain Constraints

Yu et al., '08
Del Pero et al., '12
Hedau et al., '09
Schwing et al., '12
Lee et al., '10
Man-Made Constraints
But where are the people?
People as Clutter?

People Occlude the Scene!
Humans tell a lot about geometry
Timelapse
Timelapse

Pose Detections
Estimate Functional Regions from Poses
3D Room Hypotheses From Appearance
Score 3D Room Hypotheses With Appearances + Affordances
ECCV 2012

People Watching: Human Actions as a Cue for Single View Geometry

Paper #331
<table>
<thead>
<tr>
<th>Location</th>
<th>Appearance Only</th>
<th>Appearance + People</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee et al. '09</td>
<td>66.4%</td>
<td>79.6% (+2.6%)</td>
</tr>
<tr>
<td>Hedau et al. '09</td>
<td>71.3%</td>
<td>77.0%</td>
</tr>
</tbody>
</table>

Does equivalently or better 88% of the time
Qualitative

More reasoning

Region labels

Stronger geometric constraints from domain knowledge

Physical Reasoning

Semantic and Functional Reasoning