Learning from 3D Data for Image Interpretation

Martial Hebert
Abhinav Gupta
David Fouhey, Adrien Matricon, Wajahat Hussain
• Mid-level primitives learned from image+3D can be used to transfer geometric information?
• Geometric reasoning can use this local evidence to produce a consistent geometric interpretation?
Pattern Repetition

Common patterns correspond to common geometric configurations
Pattern Repetition
Pattern Repetition
Physical/Geometric Constraints
Primitives

Visually Discriminative

Geometrically Informative

Image

Surface Normals

Saurabh Singh et al. *Discriminative Mid-Level Patches*
Geometric configurations from large-scale RGBD data.

NYU v2 Dataset (Silberman et al., 2012)
Representation

Detector

Canonical Form

Instances

10x10
Representation

Detector

Canonical Form

Instances
Learning Primitives

\[
\min_{y,w,N} R(w) + \sum_{i} c_1 y_i \Delta(N, x_i^G) + c_2 L(w, N, x_i^A, y_i)
\]

**Primitive**

\[
\begin{array}{c}
w \\
N \\
y
\end{array}
\]

**Patch**

\[
\begin{array}{c}
x_i^A \\
x_i^G
\end{array}
\]
Learning Primitives

Approach: iterative procedure
Learning Primitives

\[ N = \text{Avg}(w) \]
Learning Primitives

N

w

Cluster Instances

Patches Geometrically Dissimilar to N
Learning Primitives
Learning Primitives

Initialize \( y \) by clustering sampled patches
Inference

Sparse Transfer
Inference

Sparse Transfer
Inference

Sparse Transfer
Inference

Dense Transfer
Sample Results – Qualitative
Confidences

Most Confident Result

Least Confident Result
Cross-dataset

PETS

B3DO
Failures
<table>
<thead>
<tr>
<th></th>
<th>Summary Stats (°) (Lower Better)</th>
<th>% Good Pixels (Higher Better)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>3D Primitives</td>
<td>33.0</td>
<td>28.3</td>
</tr>
<tr>
<td>Singh et al.</td>
<td>35.0</td>
<td>32.4</td>
</tr>
<tr>
<td>Karsch et al.</td>
<td>40.8</td>
<td>37.8</td>
</tr>
<tr>
<td>Hoiem et al.</td>
<td>41.2</td>
<td>34.8</td>
</tr>
<tr>
<td>Saxena et al.</td>
<td>47.1</td>
<td>42.3</td>
</tr>
<tr>
<td>RF + Dense SIFT</td>
<td>36.0</td>
<td>33.4</td>
</tr>
</tbody>
</table>
Using geometric and physical constraints
The Story So Far (Sparse)
The Story So Far (Dense)
The Story So Far
Adding Physical/Geometric Constraints
Adding Physical/Geometric Constraints
Past Physical Constraints

Camera-in-a-box


Top-down Cuboid

Lee et al. 2010, Gupta et al. 2010, Xiao et al. 2012, etc.
Digression: Inspiration from the past....

Kanade’s Origami World, 1978
From the past....

• Kanade’s chair... (Artificial Intelligence, 1981)
Edges between surfaces

Concave ( - )

Convex ( + )
Edges between surfaces

Concave
( - )

Convex
( + )
Parameterization

$\text{vp}_1$

$\text{vp}_2$

$\text{vp}_3$
Parameterization

$vp_1$, $vp_2$, $vp_3$

Schwing 2013, Hedau 2010
Parameterization

$vp_1$

$vp_2$

$vp_3$
Parameterization
Parameterization
Parameterization
Parameterization
$\mathcal{X}_i$: is cell $i$ on?

$x \in \{0,1\}^n$
Formulation

$$\arg \max c^T x + x^T H x \quad \text{s.t.} \quad Ax \leq 1$$
$$x \in \{0, 1\}^n$$
$X_i$: is cell $i$ on?

$$\arg \max_{x \in \{0,1\}^n} c^T x + x^T H x \quad \text{s.t.} \quad Ax \leq 1$$
Unary Potentials

\( C_i : \text{should cell } i \text{ be on?} \)

\[
\arg \max_{x \in \{0,1\}^n} c^T x + x^T H x \quad \text{s.t.} \quad Ax \leq 1
\]
Binary Potentials

$H_{i,j}$: should cells $i$ and $j$ both be on?

$$\arg \max_{x} c^T x + x^T H x \quad \text{s.t.} \quad Ax \leq 1$$

$x \in \{0,1\}^n$
Binary Potentials

Convex ( + )

Concave ( - )
Binary Potentials

Convex (+)

Concave (-)
Constraints

What configurations are forbidden?

\[
\arg \max \mathbf{c}^T \mathbf{x} + \mathbf{x}^T \mathbf{H} \mathbf{x} \quad \text{s.t.} \quad \mathbf{A} \mathbf{x} \leq 1
\]

\[\mathbf{x} \in \{0,1\}^n\]
Projected 3D Primitives

Input

Ground Truth

3D Primitives

Projected 3D Primitives

Proposed
Qualitative Results

Input

Ground Truth

3D Primitives

Projected 3D Primitives

Proposed
Projected 3D Primitives

Input

Ground Truth

3D Primitives

Projected 3D Primitives

Proposed
Random Qualitative Results

3D Primitives

Proposed
## Quantitative Results

<table>
<thead>
<tr>
<th></th>
<th>Summary Stats (°)</th>
<th>% Good Pixels (Higher Better)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Proposed</td>
<td>37.5</td>
<td>17.2</td>
</tr>
<tr>
<td>3D Primitives</td>
<td>38.5</td>
<td>19.0</td>
</tr>
<tr>
<td>Hedau et al.</td>
<td>43.2</td>
<td>24.8</td>
</tr>
<tr>
<td>Lee et al.</td>
<td>47.6</td>
<td>43.4</td>
</tr>
<tr>
<td>Karsch et al.</td>
<td>46.6</td>
<td>43.0</td>
</tr>
<tr>
<td>Hoiem et al.</td>
<td>45.6</td>
<td>38.2</td>
</tr>
</tbody>
</table>

*Rank: Proposed*
Style vs. structure?

Casablanca Hotel, New York
More general environments?
KITTI Dataset: Geiger, Lenz, Urtasun, ‘12
• Large regions without surface interpretation
• Fewer linear/planar structures to anchor
• Irregular distribution of 3D training data
Discovered Primitives (Examples)
Contact points
Object surfaces + Contact points
Next:
- Better reasoning
- Semantic information
- Less structured environments
- Evaluation
- Applications

Data-Driven 3D Primitives For Single-Image Understanding, Fouhey, Gupta, Hebert, In ICCV 2013.

Unfolding an Indoor Origami World, Fouhey, Gupta, Hebert, In ECCV 2014.
- Harvested from tripadvisor.com

<table>
<thead>
<tr>
<th>Countries</th>
<th>8</th>
<th>USA, Japan, London, Germany, Canada, Australia, Thailand, Indonesia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chains</td>
<td>~ 5</td>
<td>Hilton, Marriott, Intercontinental, Sheraton, Best Western etc.</td>
</tr>
</tbody>
</table>
Sheraton Los Angeles

Meritan Apartments Sydney

Le Champlain Quebec
Project digression.....

CyberTipline

The CyberTipline® receives leads and tips regarding suspected crimes of sexual exploitation committed against children. More than 2.3 million reports of suspected child sexual exploitation have been made to the CyberTipline between 1998 and March 2014.

If you have information regarding possible child sexual exploitation, report it to the CyberTipline.

MAKE A CYBERTIPLINE REPORT

Purpose and function

The CyberTipline is operated in partnership with the FBI, Immigration and Customs Enforcement, U.S. Postal Inspection Service, U.S. Secret Service, military criminal investigative organizations, U.S. Department of Justice, Internet Crimes Against Children Task Force program, as well as other state and local law enforcement agencies. Reports to the CyberTipline are made by the
Next:
  Better reasoning
  Semantic information
  Less structured environments
  Evaluation
  Applications


*Unfolding an Indoor Origami World*, Fouhey, Gupta, Hebert, In ECCV 2014.
Results – Quantitative

The graph shows the quantitative recall for different methods. The x-axis represents recall, while the y-axis ranges from 0 to 1. The different methods and their corresponding colors are:

- Yellow: Proposed
- Pink: No Iterations
- Blue: Singh et al.
- Red: Hoiem et al.

The proposed method outperforms the others at lower recall values, but the performance gap narrows as recall increases.