Using 3D data for image interpretation and geometric reasoning

Martial Hebert
Abhinav Gupta
David Fouhey, Adrien Matricon, Wajahat Hussain
• Sparse mid-level primitives can be used to transfer geometric information?
• Can this help in detection and matching tasks?
• Geometric reasoning can use this local evidence to produce a consistent geometric interpretation?
Primitives

Visually Discriminative

Geometrically Informative

Saurabh Singh et al. Discriminative Mid-Level Patches
NYU v2 Dataset (Silberman et al., 2012)
Learning primitives
Representation

Detector

Instances

Canonical Form
Learning Primitives

Approach: iterative procedure
Inference

Sparse Transfer
Inference

Sparse Transfer
Inference

Dense Transfer
Sample Results – Qualitative
Confidence

Most Confident Result

Least Confident Result
Failures
<table>
<thead>
<tr>
<th>Method</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>
More general environments?
• Large regions without surface interpretation
• Fewer linear/planar structures to anchor
• Irregular distribution of 3D training data
Discovered Primitives (Examples)
<table>
<thead>
<tr>
<th>Method</th>
<th>Summary Stats (º)</th>
<th>% Good Pixels (º)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Lower Better)</td>
<td>(Higher Better)</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>3D Primitives</td>
<td>23.4</td>
<td>9.9</td>
</tr>
<tr>
<td>RF + Dense SIFT</td>
<td>24.2</td>
<td>16.3</td>
</tr>
</tbody>
</table>
Contact points
Object surfaces + Contact points
Failures
Digression
Style and structure
Style vs. structure?

Casablanca Hotel, New York
Meritan Apartments Sydney

Sheraton Hotels (North America)
Using geometric and physical constraints
The Story So Far
The Story So Far
Adding Physical/Geometric Constraints
Adding Physical/Geometric Constraints
Huffman 71, Clowes 71, Kanade 80, 81, Sugihara 86, Malik 87, etc.
Edges between surfaces

Concave ( - )

Convex ( + )
Parameterization

$vp_1$

$vp_2$

$vp_3$
Parameterization

\( \text{vp}_1 \)

\( \text{vp}_2 \)

\( \text{vp}_3 \)
Parameterization

vp_1

vp_2

vp_3
Parameterization
Parameterization
Parameterization
$X_i$: is cell $i$ on?
Unary terms

Should cell $i$ be on?
Binary Potentials
Binary terms
Binary terms
Binary terms
Constraints
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>Summary Stats ($^0$)</th>
<th>% Good Pixels (Higher Better)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
</tr>
<tr>
<td>Proposed</td>
<td>37.5</td>
</tr>
<tr>
<td>3D Primitives</td>
<td>38.5</td>
</tr>
<tr>
<td>Hedau et al.</td>
<td>43.2</td>
</tr>
<tr>
<td>Lee et al.</td>
<td>47.6</td>
</tr>
<tr>
<td>Karsch et al.</td>
<td>46.6</td>
</tr>
<tr>
<td>Hoiem et al.</td>
<td>45.6</td>
</tr>
</tbody>
</table>
Now:

Better reasoning
Semantic information
Less structured environments
Coarse-to-fine depth
Martial Hebert
Abhinav Gupta
David Fouhey, Adrien Matricon, Wajahat Hussain

ONR MURI
NDSEG
Bosch R&D