Can Similar Scenes help Surface Layout Estimation?

Santosh K. Divvala
Alexei A. Efros
Martial Hebert

Carnegie Mellon University
The Robotics Institute
Single-view Scene Understanding

- Face of a Building 8m away at 45°
- 10m away at -45°
- 15m away at 30°
- Tree
- Sky
- Ground
Learning-based Methods

Geometric Context [Hoiem et al. ‘05,’07]

Scene Structure [Saxena et al. ‘05,’07]

Stage Classification [Nedovic et al. ‘07]
Labeling Image Data

Tree or Building?
Alternate Solution(s)?

Yes. Hire a fresh graduate student!
Using Unlabeled Data

Sparse Labeled Data

Image with labels

Lots of Unlabeled Data

Labeled Data

Output

Classifier

Input

?
Color Legend

- Sky
- **Vertical** Left (←), Center (↑), Right (→), Porous (○), Solid (X)
- Support

[http://make3d.stanford.edu]
Similarity

Scene

Similar Scenes
Works using Similar Images

[Hays & Efros '07]

Gray-scale Input  Gray-level Siblings  High-resolution color siblings  Average Color  Average Colorization

[torralba et al. '07]
Overview

Image

Labeled Data

Multiple Segmentations

Surface Estimates

Confidence Map

Geometric Context (IJCV’07)

Unlabeled Data

Scene Matches

1. Prior confidence map for improved surface classification

2. Average image features for improved surface segmentations

This Paper
Retrieving Scene Matches

Thanks to James Hays
Generating Scene Matches
Clustering Scene Matches

Scene Match
Geometric Context Output

Sky
Left
Center
Right

Porous
Solid
Support
Clustered Scene Matches

After Clustering
Computing Average Scene Prior

Average Scene Prior

Confidence Maps of the Clustered Matches
What does the average scene prior convey?

1. Single Large Segment
2. Belongs to class ‘Vertical’

Incorporate as feature
Using Average Scene Prior

Top few scene matches and their Geometric Context outputs

New Result

Ground-Truth
<table>
<thead>
<tr>
<th>Input Image</th>
<th>Ground-Truth</th>
<th>Hoiem <em>et al.</em></th>
<th>New Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.jpg" alt="Image" /></td>
<td><img src="image2.jpg" alt="Image" /></td>
<td><img src="image3.jpg" alt="Image" /></td>
<td><img src="image4.jpg" alt="Image" /></td>
</tr>
<tr>
<td><img src="image5.jpg" alt="Image" /></td>
<td><img src="image6.jpg" alt="Image" /></td>
<td><img src="image7.jpg" alt="Image" /></td>
<td><img src="image8.jpg" alt="Image" /></td>
</tr>
</tbody>
</table>
Overview

Image

Multiple Segmentations

Surface Estimates

Confidence Map

Geometric Context (IJCV’07)

Labeled Data

1. Scene prior for improved surface classification

2. Average image features for improved surface segmentations

Unlabeled Data

Scene Matches

This Paper
Good Segmentation is Important

[Malisiewicz & Efros '07]
Average Scene Statistics

Input Image

Scene Matches

Segment 1 (sky)
Segment 2 (bldg)
Segment 3 (grnd)

Average Scene Statistic

Average Image
Average Scene Features

- Color Features (#14)
- Texture Features (#30)
- Location Features (#8)
Segmentations using Average Features
Segmentations using Average Features

Shore Scene

Hoiem et al.
Segmentations using Average Features

Shore Scene

Top few scene matches

Segmentations from average features

Segmentations from image features
Results

<table>
<thead>
<tr>
<th>Input Image</th>
<th>Ground-Truth</th>
<th>Hoiem et al.</th>
<th>New Result</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
</tr>
<tr>
<td>Input Image</td>
<td>Ground-Truth</td>
<td>Hoiem et al.</td>
<td>New Result</td>
</tr>
<tr>
<td>-------------</td>
<td>--------------</td>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td><img src="image1.png" alt="Input Image" /></td>
<td><img src="image2.png" alt="Ground-Truth" /></td>
<td><img src="image3.png" alt="Hoiem et al." /></td>
<td><img src="image4.png" alt="New Result" /></td>
</tr>
<tr>
<td><img src="image5.png" alt="Input Image" /></td>
<td><img src="image6.png" alt="Ground-Truth" /></td>
<td><img src="image7.png" alt="Hoiem et al." /></td>
<td><img src="image8.png" alt="New Result" /></td>
</tr>
</tbody>
</table>
Quantitative Evaluation

- Input: Geometric Context Imageset (300)
- Scene Matches: Flickr (6,000,000)
## Performance

<table>
<thead>
<tr>
<th></th>
<th>Sub-Class Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Features (Hoiem et al.)</td>
<td>60.5%</td>
</tr>
<tr>
<td>Average Scene Prior</td>
<td>52.3%</td>
</tr>
<tr>
<td>Average Scene Prior + Image Features</td>
<td>63.2%*</td>
</tr>
<tr>
<td>Average Feature Segmentations</td>
<td>61.5%*</td>
</tr>
<tr>
<td>Both the above</td>
<td>63.6%*</td>
</tr>
</tbody>
</table>

- **Input Image**
- **Ground Truth**
- **Average Scene Prior**
## Performance

<table>
<thead>
<tr>
<th></th>
<th>Sub-Class Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Features</td>
<td>60.5%</td>
</tr>
<tr>
<td>Average Scene Prior</td>
<td>52.3%</td>
</tr>
<tr>
<td>Average Scene Prior + Image Features</td>
<td>63.2%*</td>
</tr>
<tr>
<td>Average Feature Segmentations</td>
<td>61.5%*</td>
</tr>
<tr>
<td>Both the above</td>
<td>63.6%*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(use subset of data)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ramalingam etal (Tuesday – Poster1A)</td>
<td></td>
</tr>
<tr>
<td>Hoiem etal (Wednesday – Oral2A)</td>
<td>61.8%</td>
</tr>
</tbody>
</table>

(* p-value > 0.05)
Qualitative Observations

Canonical Scenes -> Good Scene Matches -> Improved Results

- Alley
- Field
- Building
- Cliff

Tougher Scenes

- City
- Street
- Outdoor
Conclusion

• Scene understanding algorithms can utilize unlabeled data

• Proposed specific ideas based on ‘similarity’

• Improved Geometric Context system
Acknowledgement

• James Hays - Generating scene matches

• Derek Hoeim – Geometric Context code

• Support from NSF
Thank You

Input Image  Ground-Truth  Hoiem et al.  New Result
Geometric Context: Cues

- Color
- Texture
- Location
- Perspective
# Geometric Context: Features

<table>
<thead>
<tr>
<th>Feature Descriptions</th>
<th>Num</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Color</strong></td>
<td></td>
</tr>
<tr>
<td>C1. RGB values: mean</td>
<td>16</td>
</tr>
<tr>
<td>C2. HSV values: C1 in HSV space</td>
<td>3</td>
</tr>
<tr>
<td>C3. Hue: histogram (5 bins) and entropy</td>
<td>6</td>
</tr>
<tr>
<td>C4. Saturation: histogram (3 bins) and entropy</td>
<td>4</td>
</tr>
<tr>
<td><strong>Texture</strong></td>
<td></td>
</tr>
<tr>
<td>T1. DOOG filters: mean abs response of 12 filters</td>
<td>15</td>
</tr>
<tr>
<td>T2. DOOG stats: mean of variables in T1</td>
<td>12</td>
</tr>
<tr>
<td>T3. DOOG stats: argmax of variables in T1</td>
<td>1</td>
</tr>
<tr>
<td>T4. DOOG stats: (max - median) of variables in T1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Location and Shape</strong></td>
<td></td>
</tr>
<tr>
<td>L1. Location: normalized x and y, mean</td>
<td>12</td>
</tr>
<tr>
<td>L2. Location: norm. x and y, 10&lt;sup&gt;th&lt;/sup&gt; and 90&lt;sup&gt;th&lt;/sup&gt; petl</td>
<td>2</td>
</tr>
<tr>
<td>L3. Location: norm. y wrt horizon, 10&lt;sup&gt;th&lt;/sup&gt;, 90&lt;sup&gt;th&lt;/sup&gt; petl</td>
<td>4</td>
</tr>
<tr>
<td>L4. Shape: number of superpixels in region</td>
<td>2</td>
</tr>
<tr>
<td>L5. Shape: number of sides of convex hull</td>
<td>1</td>
</tr>
<tr>
<td>L6. Shape: [\text{num pixels/area(convex hull)}]</td>
<td>1</td>
</tr>
<tr>
<td>L7. Shape: whether the region is contiguous (\in{0, 1})</td>
<td>1</td>
</tr>
<tr>
<td><strong>Geometry</strong></td>
<td></td>
</tr>
<tr>
<td>G1. Texture gradient: x and y &quot;edginess&quot; (T2) center</td>
<td>35</td>
</tr>
<tr>
<td>G2. Long lines: total number in region</td>
<td>2</td>
</tr>
<tr>
<td>G3. Long lines: % of nearly parallel pairs of lines</td>
<td>1</td>
</tr>
<tr>
<td>G4. Line int: hist. over 12 orientations, entropy</td>
<td>13</td>
</tr>
<tr>
<td>G5. Line int: % right of center</td>
<td>1</td>
</tr>
<tr>
<td>G6. Line int: % above center</td>
<td>1</td>
</tr>
<tr>
<td>G7. Line int: % far from center at 8 orientations</td>
<td>8</td>
</tr>
<tr>
<td>G8. Line int: % very far from center at 8 orientations</td>
<td>8</td>
</tr>
</tbody>
</table>
Using Average Scene Prior

Top few scene matches and their Geometric Context outputs

New Result

Ground-Truth
Using Average Scene Prior

Input Image | Ground-Truth | Hoiem et al. | New Result
Performance with varying neighbors
Relative Performance Improvement

![Histogram showing the relative performance improvement in accuracy. The x-axis represents the relative improvement in accuracy, ranging from -1 to 1. The y-axis represents the number of images. The histogram peaks around a relative improvement of 0, indicating a significant number of images have improved their accuracy by this amount.](image-url)
Co-Training

Color | Location | Perspective | Texture
--- | --- | --- | ---
![Image](image1.png) | ![Image](image2.png) | ![Image](image3.png) | ![Image](image4.png)

Average Scene Prior | Hoiem et al.
Thank You

Input Image | Ground-Truth | Hoiem et al. | New Result