Qualitative 3D Surface Reconstruction From Images

Alex Levinshtein\textsuperscript{1}, Cristian Sminchisescu\textsuperscript{2} and Sven Dickinson\textsuperscript{1}, \textsuperscript{1}University of Toronto, \textsuperscript{2}University of Bonn

**Goal: Qualitative 3D from a single image**

- **Input (image)**
- **Output (piecewise surface labeling)**

**Approach**
Learn a mapping from images to qualitative surface shape

**Motivation**
- A mid-level object description
- Geometric coarser than exact reconstruction, not object specific, however
- Useful for indexing, even if sparse
- Useful for categorical recognition

**Execution**
Train a Conditional Random Field (CRF) to recover piecewise labeling from image superpixel graph, based on (2d,3d) training set

### Related Work

**Geometric Labels**
- 3D Model
  - General
  - Vertical
  - Concave
  - Convex
  - Sagittal

**Novel view**
- Saxena et al., 3DRR-07
- Hohem et al., SIGGRAPH 2005

**Reconstructed Mesh**

**Input**

**3b. CRF inference**

- Use Loopy Belief propagation for inference
- At every iteration update messages in parallel
- Use damping for stability

\[
\mu_{\text{new}} = (1-\alpha)\mu_{\text{old}} + \alpha\mu_{\text{ad}}
\]

**4. Results**

Test on two datasets (Minolta and K2T) containing pairs of optical and range images

**Qualitative Results**
- Image
- Ground Truth
- Inferred labels

**Quantitative Results**

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minolta</td>
<td>0.6</td>
<td>0.48</td>
</tr>
<tr>
<td>K2T</td>
<td>0.48</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Precision on 6-label problem (all saddles merged into one class)

**3a. Bethe Free Energy Approximations for CRF**

- Use Conditional Maximum likelihood learning
- Compute log-likelihood of the data under the model
  - Use Bethe Free Energy to approximate log Z

\[
\log Z = -F_Z
\]

**2. Extract features**

- Each superpixel is associated feature vector \( f_i \)
- Features capture
  - Contour strength and curvature
  - Moments of intensity inside superpixel, local jets, etc

**1. Construct superpixel graph**

- Over-segment the image
- Construct a graph whose segments are nodes and edges are neighborhood relations between nodes

**3. Train CRF**

- Find the maximum likelihood parameters of the CRF

**Input (image)**

**Output (piecewise surface labeling)**

**Convex**

**Concave**

**Cylindrical**

**1. Superpixel graph**

- Over-segment the image using Normalized Cuts
- Each superpixel becomes node in graph
- Neighboring superpixels (those with common boundary) share an edge

**2. Feature Design**

- 36 dimensional vector \( f \) encodes both contour and internal superpixel appearance information
- Kernelize by computing covariance features for each superpixel
  - Estimate \( F, F^T \)
  - Vectorize lower triangular the matrix as feature set

**3. CRF formulation**

\[
p(X | F, w, v) = \frac{1}{Z(w,v)} \exp \left( \sum w_i F_i + \sum v_i (i,j) \right)
\]

**CRF parameters:**
- \( w \) = feature vector for superpixel \( i \)
- \( F_i \) = feature vector for superpixel \( i \) (after the kernelization)
- \( K \) = number of possible labels (8 for HK encoding)
- \( S \) = number of superpixels

**Derivative of the partition function involves intractable expectation**

- Use loopy belief propagation to approximate necessary marginals

**Overview of the Method**

- Use damping for stability
- is associated
- x
- Useful for categorical recognition
- number of possible
- K
- are the surface (local) principal curvatures:

\[
H = \frac{1}{2} \left( k_1 + k_2 \right) \quad K = k_1 k_2
\]

**Input (image)**

**Goal: Qualitative 3D from an image**

**Related Work**

**Geometric Labels**
- 3D Model
  - General
  - Vertical
  - Concave
  - Convex
  - Sagittal

**Novel view**
- Saxena et al., 3DRR-07
- Hohem et al., SIGGRAPH 2005

**Reconstructed Mesh**

**Input**

**3b. CRF inference**

- Use Loopy Belief propagation for inference
- At every iteration update messages in parallel
- Use damping for stability

\[
\mu_{\text{new}} = (1-\alpha)\mu_{\text{old}} + \alpha\mu_{\text{ad}}
\]

**4. Results**

Test on two datasets (Minolta and K2T) containing pairs of optical and range images

**Qualitative Results**
- Image
- Ground Truth
- Inferred labels

**Quantitative Results**

<table>
<thead>
<tr>
<th></th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minolta</td>
<td>0.6</td>
<td>0.48</td>
</tr>
<tr>
<td>K2T</td>
<td>0.48</td>
<td>0.32</td>
</tr>
</tbody>
</table>

Precision on 6-label problem (all saddles merged into one class)