Recognition of Deformable Object Category and Pose

Columbia University

Yinxiao Li
Chih-Fan Chen
Peter Allen
What is a deformable object (DO)?
What is a deformable object (DO)?
What is a deformable object (DO)?

What is a deformable object (DO)?

http://www.kidnesia.com/Kidnesia/Potret-Negeriku/Flona/Fauna/Cacing-Tanah
Why do robots need to recognize DO?

http://www.cs.berkeley.edu/~pabbeel/
http://www2.informatik.uni-freiburg.de/~stachnis/pdf/frank10rssws.pdf
Our goal

http://www.willowgarage.com
Challenge – Find Category & Pose
Outline

• Entire Pipeline of Manipulation of DO
• Training and Testing Flow
  – Defining Deformable Poses
  – Generating Training Exemplars
  – Estimating Poses of DO
  – Two-Layer Classifier
• Experimental Results
• Conclusion
Pipeline for Manipulation of DO
Pipeline for Manipulation of DO
Pipeline for Manipulation of DO
Pipeline for Manipulation of DO
Pipeline for Manipulation of DO
Pipeline for Manipulation of DO

Recognition and identification of the garment
Entire Training and Testing Flow
Entire Training and Testing Flow

Training:
- Off-line Simulation
- 90-Camera system
- Depth Images From The Mesh Models

Testing:
- The Kinect Sensor
- Crop Image
- Depth Image
- Images From The Kinect

VLFeat
Sift Feature Vectors

Sparse Coding
1 layer: RBF Kernel
2 layer: Linear Kernel
SVM Classifiers

Pruning Labels
Output Labels
Entire Training and Testing Flow

Training

90-Camera system
Off-line Simulation

Sift Feature Vectors

Sift Feature Vectors

Testing

The Kinect Sensor
Crop Image
Depth Image

Code Book

1 layer: RBF Kernel
2 layer: Linear Kernel
SVM Classifiers

Pruning Labels
Output Labels

Yinxiao Li
Entire Training and Testing Flow

Training:
- Off-line Simulation
- 90-Ca system

Testing:
- The Kinect Sensor
- Crx Imu
- Def

Encode
Sparse Coding
1 layer
Layer: RBF Kernel
2 layer
Layer: Linear Kernel
SVM Classifiers
Pruning Labels
Output Labels

Code Book

Yinxiao Li
Defining Deformable Poses

- 20-50 points on each garment
- Color coding label pads on real garment
Generating Training Exemplars

- Virtual cameras in Maya (cloth simulator)
- 90-Camera System on a geodesic dome
Generating Training Exemplars

- Virtual cameras in Maya (cloth simulator)
- 90-Camera System on a geodesic dome
Generating Training Exemplars

- Crop image in terms of the object
- Dense SIFT feature over the entire image
Learning feature signature

- Sparse coding

\[
\min_{\mathbf{w}, \mathbf{V}} \sum_{i=1}^{N} \| \mathbf{X}_i - \mathbf{w}_i \mathbf{V} \|^2 + \lambda \| \mathbf{w}_i \|
\]

- Max pooling

\[ r_j = \max\{|w'_{1j}|, |w'_{2j}|, \ldots, |w'_{Nj}|\} \]

- SVM
Two-Layer Classifier

- First Layer: Category classification (SVM₁)
- Second Layer: Pose identification (SVM₂)

Diagram:
- Input Depth Images
- Feature Vector
- Garment Types: Sweater, Jean, Skirt
- Grasping Point Labels: 7, 18, 36, 1, 10, 33, 2, 5, 15
Test using simulation data
Test using simulation data
Test using simulation data

<table>
<thead>
<tr>
<th>Height Group</th>
<th># of Cam</th>
<th>Sweater Rank 1</th>
<th>Jean Rank 1</th>
<th>Short Pant Rank 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>9</td>
<td>6.06</td>
<td>6.12</td>
<td>7.19</td>
</tr>
<tr>
<td>Group 2</td>
<td>18</td>
<td>4.93</td>
<td>5.12</td>
<td>5.67</td>
</tr>
<tr>
<td>Group 3</td>
<td>36</td>
<td><strong>4.52</strong></td>
<td><strong>4.44</strong></td>
<td><strong>5.48</strong></td>
</tr>
<tr>
<td>Group 4</td>
<td>18</td>
<td>4.84</td>
<td>5.06</td>
<td>6.29</td>
</tr>
<tr>
<td>Group 5</td>
<td>9</td>
<td>5.38</td>
<td>6.05</td>
<td>7.29</td>
</tr>
<tr>
<td>Group 2, 3, 4</td>
<td>72</td>
<td>5.02</td>
<td>5.09</td>
<td>5.81</td>
</tr>
<tr>
<td>All</td>
<td>90</td>
<td>5.79</td>
<td>5.73</td>
<td>6.76</td>
</tr>
</tbody>
</table>
Test using depth image from the Kinect

Color Image

Depth image

Ground truth from database

Predicted result from database

<table>
<thead>
<tr>
<th>38(38)</th>
<th>33(36)</th>
<th>40(39)</th>
<th>22(24)</th>
<th>4(40)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0 cm</td>
<td>4.6 cm</td>
<td>3.2 cm</td>
<td>5.9 cm</td>
<td>22.6 cm</td>
</tr>
</tbody>
</table>
Test using depth image from the Kinect

Color Image

Depth image

Ground truth from database

Predicted result from database

33 (36)
4.6 cm
Test using depth image from the Kinect

Color Image

Depth image

Ground truth from database

Predicted result from database

Ground truth grasping point

Predicted grasping point

Distance between the two points

Distance: 4.6 cm

Grasping points:
- Ground truth: (33, 36)
- Predicted: (33, 36)

Table:

<table>
<thead>
<tr>
<th>Distance</th>
<th>Grasping Point</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.6 cm</td>
<td>(33, 36)</td>
</tr>
<tr>
<td>38 cm</td>
<td>(38, 38)</td>
</tr>
<tr>
<td>33 cm</td>
<td>(33, 36)</td>
</tr>
<tr>
<td>40 cm</td>
<td>(40, 39)</td>
</tr>
<tr>
<td>22 cm</td>
<td>(22, 24)</td>
</tr>
<tr>
<td>4 cm</td>
<td>(4, 40)</td>
</tr>
</tbody>
</table>

Yinxiao Li
Experimental Results

• Test using depth image from the Kinect

<table>
<thead>
<tr>
<th>Height Group</th>
<th>Garment Categories</th>
<th>Sweater</th>
<th>Jean</th>
<th>Short Pant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 3</td>
<td></td>
<td>75.79%</td>
<td>63.33%</td>
<td>84.71 %</td>
</tr>
<tr>
<td>Group 2, 3, 4</td>
<td></td>
<td>79.61%</td>
<td>62.83%</td>
<td>90.76 %</td>
</tr>
<tr>
<td>All</td>
<td></td>
<td>73.77%</td>
<td>72.73%</td>
<td>91.40 %</td>
</tr>
</tbody>
</table>
Accuracy plot for sweater

Max distance between two points is 65 cm

15 cm offset is good enough for a robot to manipulate a garment!
Implementation on a robot -- Video
Implementation on a robot

1. Grasping
2. Picking up
3. Rotation and Recognition

Sweater: avg distance = 7.8 cm
Pants: avg distance = 10.2 cm
Conclusion

• A framework of recognizing deformable clothing object category and poses
• **Data-Driven** approach: off-line simulation, on-line recognition
• Experiments on both simulation data and real depth images—high recognition rate / accuracy

Future Work

• **Kinect fusion** algorithm to reconstruct
• Employing **color and texture** features
• Publish a **database** of deformable object with simulation
• Further tasks of manipulation – **regrasping**, placing flat, folding clothing
Acknowledgement

• Columbia Computer Vision and Graphics Center (CVGC)
  – Eitan Grinspun, Austin Reiter, Jon Weisz, Akash Grag, Thomas Berg

• Anonymous reviewers
Recognition of Deformable Object Category and Pose

Q&A

Columbia University

Yinxiao Li, Chih-Fan Chen, and Peter Allen