Using Microsoft Kinect Sensor in Our Research

Hao Zhang
Distributed Intelligence Laboratory
Dept. of Electrical Engineering and Computer Science
University of Tennessee, Knoxville

Sep. 20, 2011
• Introduction of Kinect Sensor
• Using Kinect in Our Research
• Introduction of Kinect Sensor
• Using Kinect in Our Research
• What is a Kinect sensor?
  – Kinect is a motion sensing device by Microsoft for the Xbox 360 video game console.
  – Kinect contains a RGB camera, a depth sensor, multi-array microphones, and a motorized tilt.
• How does a Kinect sense depth?
  – The IR emitter projects an irregular pattern of IR dots of varying intensities.
  – The Depth Camera reconstructs a depth image by recognizing the distortion in this pattern.
What’s the accuracy of a Kinect sensor?

- Data Stream
  - 640X480, 320X240 in Linux and Mac
  - 1024X768, 640X480, 320X240 in Windows 7
  - 30 frames/sec

- Depth Camera
  - Field of View
    - Horizontal: 58°, Vertical: 45°, Diagonal: 70°
  - Spatial X/Y resolution: 3mm
  - Depth Z resolution: 1cm
  - Operation range: 0.8m - 3.5m

- Physical Tilt Range: ±27 degrees
• Introduction of Kinect Sensor
• Using Kinect in Our Research
• Why do we choose Kinect?
  – Powerful
    • Capable of acquiring color, depth, and audio information
  – Not expensive
    • $150 each (a sensor and power supply)
  – Accessible
    • Available at game stores, computer stores, and supermarkets
  – Easy to setup and use
• First step toward making Kinect work: **Install A Driver for Kinect Sensor and related dependencies**

  - Kinect for Windows SDK
    - Support Windows 7 only
  
  - OpenKinect
    - OpenNI Kinect
    - Libfreenect
      - Supporting Windows, Mac and Linux
      - Combined in ROS
• Kinect in My Research
  – Human activity recognition: automated detection of ongoing events from visual data containing movements with particular semantic meanings
• Perception using Kinect (Feature extraction)
  – 3D centroid trajectory
  – 3D shape history
  – Motion sequence of 3D human models

3D Trajectory (O. Brdiczka, 09)

3D Human Models
(J. Y. Sung, PAIR11) & (S. Knoop, ICRA06)

3D Shape Info. (P. Yan, CVPR08)
• **4D Local Spatio-Temporal (LST) Features**
  
  – A LST feature can represent local texture and motion variations regardless of global human appearance and activity (locality assumption)
  
  – Visual data and human activity can be presented as a bag of LST features (representativeness assumption)
• Installation: on a Pioneer 3DX mobile robot
• Preprocessing of Kinect Data

1. Acquire Depth Image
2. Resize to 320 x 240
3. Erode and Dilate
4. Fill Holes
5. Depth Data

6. Acquire Color Image
7. Resize to 320 x 240
8. Convert to Intensity Image
9. Equalize Histogram
10. Intensity Data
• Feature Extraction
• **Activity Dataset**
  – 6 types of human activities
  – 33 samples for each activity
  – 2 ~ 4 seconds of each sample
  – Office and home environments

http://www.youtube.com/watch?v=ZYGmQYNvfnA  
http://www.youtube.com/watch?v=puhG5gty0XA
• Test Results
  – 4D-LST feature outperforms the features using only intensity or depth information
  – Depth information is more important than the intensity information for our database

<table>
<thead>
<tr>
<th></th>
<th>Lift</th>
<th>Remove</th>
<th>Wave</th>
<th>Push</th>
<th>Walk</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lift</td>
<td>.91</td>
<td>.09</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Remove</td>
<td>.14</td>
<td>.86</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Wave</td>
<td>.00</td>
<td>.00</td>
<td>.95</td>
<td>.00</td>
<td>.00</td>
<td>.05</td>
</tr>
<tr>
<td>Push</td>
<td>.05</td>
<td>.05</td>
<td>.00</td>
<td>.90</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Walk</td>
<td>.04</td>
<td>.01</td>
<td>.01</td>
<td>.03</td>
<td>.92</td>
<td>.00</td>
</tr>
<tr>
<td>Signal</td>
<td>.00</td>
<td>.00</td>
<td>.05</td>
<td>.00</td>
<td>.00</td>
<td>.95</td>
</tr>
</tbody>
</table>

Intensity & Depth Data  
(Average accuracy = 91.50%)

<table>
<thead>
<tr>
<th></th>
<th>Lift</th>
<th>Remove</th>
<th>Wave</th>
<th>Push</th>
<th>Walk</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lift</td>
<td>.95</td>
<td>.05</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Remove</td>
<td>.10</td>
<td>.85</td>
<td>.05</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Wave</td>
<td>.00</td>
<td>.00</td>
<td>.88</td>
<td>.00</td>
<td>.00</td>
<td>.12</td>
</tr>
<tr>
<td>Push</td>
<td>.11</td>
<td>.09</td>
<td>.79</td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td>Walk</td>
<td>.02</td>
<td>.09</td>
<td>.08</td>
<td>.07</td>
<td>.74</td>
<td>.00</td>
</tr>
<tr>
<td>Signal</td>
<td>.00</td>
<td>.00</td>
<td>.08</td>
<td>.00</td>
<td>.00</td>
<td>.92</td>
</tr>
</tbody>
</table>

Depth Data Only  
(Average accuracy = 85.50%)

<table>
<thead>
<tr>
<th></th>
<th>Lift</th>
<th>Remove</th>
<th>Wave</th>
<th>Push</th>
<th>Walk</th>
<th>Signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lift</td>
<td>.77</td>
<td>.23</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Remove</td>
<td>.23</td>
<td>.77</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Wave</td>
<td>.00</td>
<td>.00</td>
<td>.82</td>
<td>.00</td>
<td>.00</td>
<td>.18</td>
</tr>
<tr>
<td>Push</td>
<td>.14</td>
<td>.11</td>
<td>.69</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Walk</td>
<td>.05</td>
<td>.02</td>
<td>.10</td>
<td>.09</td>
<td>.68</td>
<td>.05</td>
</tr>
<tr>
<td>Signal</td>
<td>.00</td>
<td>.00</td>
<td>.05</td>
<td>.00</td>
<td>.02</td>
<td>.93</td>
</tr>
</tbody>
</table>

Intensity Data Only  
(Average accuracy = 77.67%)
Use Kinect in Your Robotics Projects

- Humanoid Robot Control and Interaction
  http://www.youtube.com/watch?v=GdeplXZTJsw

- Human Tracking and Following
  http://www.youtube.com/watch?v=3Z56JV9g6y4

- Simultaneous Localization and Mapping
  http://www.youtube.com/watch?v=XejNctt2Fcs
Thank you!