Real-Time Hand Pose Recognition using Depth Sensors combined with Spherical Shape Model Descriptor

Oscar Lopes
Advisors: Sergio Escalera
Jordi Gonzàlez

September 27, 2012
### Motivation

Real-Time Hand Pose Recognition using Depth Sensors and Shape Model Descriptor

Oscar Lopes (C)2012

<table>
<thead>
<tr>
<th>System Overview</th>
<th>Hand Detector</th>
<th>Dataset</th>
<th>Descriptor</th>
<th>Results</th>
<th>Conclusions</th>
</tr>
</thead>
</table>

**Overview:**

overview: g-speak

oblong industries
Real-Time Hand Pose Recognition using Depth Sensors and Shape Model Descriptor
Oscar Lopes (C)2012
• First the closest point is searched
  – Perform **a radius search** of 10cm, until the depth starts to rapidly increase.
  – If less **than 100 points** are obtained, then the sample is discard.
• Repeat the same procedure to find the second hand: the distance to camera must be within 30 cm of the first hand's closest distance and must more than 20 cm from the center of the first hand.
• No hand pose dataset is publicly available.
• New point cloud hand pose dataset must be created!
• The dataset was created using and adaptation of the hand detector.
• Includes **6 classes with 2000 samples** (1000 per hand):
  – Each class includes both hands.
  – High hand orientation variability.
• Plus a *No-Pose* class.
• Current state-of-art point cloud descriptors (e.g. PFH) have great computation overhead: $O(N^2)$.
• The design of a novel descriptor is necessary!
• Circular Blurred Shape Model Descriptor [Escalera et al]
  – Very good discriminative power
  – Low computational requirements

• Novel *Spherical Blurred Shape Model* Descriptor (SBSM)

\[ P = \{ p_i \mid p_i \in \mathbb{R}^3 \} \]

\[ S_R = \frac{S_{\text{Radius}}}{N_L} \]
\[ S_\theta = \frac{2\pi}{N_\theta} \]
\[ S_\phi = \frac{2\pi}{N_\phi} \]

\[ N_L \text{ number radial layers} \]
\[ N_\theta \text{ number of } \theta \text{ angular divisions} \]
\[ N_\phi \text{ number of } \phi \text{ angular divisions} \]

\[ B \text{ the ordered set of bins for the spherical description of } P^* \]

\[ b_{\{i,j,k\}} \text{ the centroid of the section } b_{\{i,j,k\}} \in B \]

\[ W_n = 0, n \in \{1, \ldots, N_L N_\phi N_\phi\} \]

```
foreach \( p_n \in P^* \) do
    \( b_x : b_x \in B, p_n \subset b_x \)
    \( W(b_x) = 1 \)
    foreach \( b_{i,j,k} \in N(b_x) \) do
        \( d_{i,j,k} = d(b_{i,j,k}, p_n) = ||p_n - b_x^*|| \)
        \( W(b_{i,j,k}) = W(b_{i,j,k}) + \frac{1}{d_{i,j,k}} \)
    end
end
```

Normalize the vector \( W \)

\[ \frac{W_i}{\#P}, i \in \{1, \ldots, N_L N_\phi N_\phi\} \]
Test settings:
- Descriptor testing considered
  \[
  N_L = \{2, 4, 8\}
  \]
  \[
  N_\theta = N_\phi = \{2, 4, 8, 16\}
  \]
- The classification was performed as multiclass one-versus-one, using the libSVM framework.
- Each combination pair was executed 10 times for cross-validation test.
  - Each execution considered 70% train data of each dataset class samples (randomly picked).
  - Every test run comprises a cross-validation of the train data for fine tune C-SVM (RBF Kernel) parameters: C and \( \gamma \).
- Previous settings were considered in two descriptor modalities:
  - Weight Propagation.
  - Zoning.
# Real-Time Hand Pose Recognition using Depth Sensors and Shape Model Descriptor

Oscar Lopes (C)2012

<table>
<thead>
<tr>
<th>Descriptor Configuration</th>
<th>Average Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Layers</td>
<td>Angles</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
</tr>
</tbody>
</table>
Multiple Hand Pose Recognition Prototype

Real-Time Hand Pose Recognition using Depth Sensors and Shape Model Descriptor
Oscar Lopes (C)2012
The proposed system achieves the proposed design goal of a real-time hand pose recognition, with an end-to-end performance of 14fps.

The SBSM descriptor is crucial for the results obtained:
- High discriminative power for hand pose point cloud.
- Small computational overhead.
- Slight advantage of blurring aspect versus Zoning, encourage further studies.

As future work...
- Creation of a more difficult **multi-user dataset**.
- Implement the descriptor algorithm using the **GPU** (Graphics processing unit) for a performance boost.
- Include a pre-description phase, for a per-pixel classification using **Random Forest** classifier, in order to perform **Label Blurring**, to increase the robustness of the overall pipeline.
Thank You!

Oscar Lopes
(oscar.pino.lopes@gmail.com)