In-air Gestures Around Unmodified Mobile Devices

Jie Song¹, Gabor Sörös¹, Fabrizio Pece¹, Sean Ryan Fanello², Shahram Izadi², Cem Keskin², Otmar Hilliges¹

¹ETH Zurich  ²Microsoft Research
Issues with Touch Input
Complement Touch with Gestures
Problem Statement

INPUT

SEGMENTATION

LABELED OUTPUT
Randomized Decision Forest

Body Pose Estimation - Shotton et al. *CVPR`11*

Hand Pose Estimation - Keskin et al. *ECCV`12*

Finger Part Classification - Kim et al. *CHI`14*

Motion Gesture Recognition - Taylor et al. *CHI`14*
Randomized Decision Tree

Input

Output

"Class A"
Runtime Efficiency vs Memory Footprint

\[ H = \log(n) \]

\[ N = \exp(H) \]

\[ 16 = 2^{(5-1)} \]
\[ \ldots = \ldots \]
\[ 2^{(30-1)} = 2^{6870912} = 4.3 \text{ GB} \]

16 bytes
Gesture Recognition with Depth Camera

[Hilliges et al. CHI’12]  [Kim et al. UIST’12]  [Wang et al., UIST ‘11]

[Harrison et al. UIST’11]  [Oikonomidis et al. UIST’11]  [Keskin et al. CPVR’12]
The Pipeline

Hand Segmentation

- Raw data
- Skin detection
- Connected Components

Coarse Depth Classification

- Too close
- Active range
- Too Far
Depth Classification Forest (DCF)

COARSE DEPTH ESTIMATION
The Pipeline

Hand Segmentation
- Raw data
- Skin detection
- Connected Components

Coarse Depth Classification
- Too close
- Active range
- Too Far

Shape Classification

The Pipeline

Hand Segmentation
- Raw data
- Skin detection
- Connected Components

Coarse Depth Classification
- Too close
- Active range
- Too Far

Shape Classification
Shape Classification Forest (SCF)
The Pipeline

Hand Segmentation
- Raw data
- Skin detection
- Connected Components

Coarse Depth Classification
- Too close
- Active range
- Too Far

Shape Classification

Part Classification
Part Classification Forest (PCF)
DCF+SCF vs SCF Only

<table>
<thead>
<tr>
<th>Depth</th>
<th>SCF</th>
<th>DCF+SCF</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.471</td>
<td>0.826</td>
</tr>
<tr>
<td>16</td>
<td>0.471</td>
<td>0.834</td>
</tr>
<tr>
<td>17</td>
<td>0.48</td>
<td>0.84</td>
</tr>
<tr>
<td>18</td>
<td>0.494</td>
<td>0.851</td>
</tr>
<tr>
<td>19</td>
<td>0.494</td>
<td>0.854</td>
</tr>
<tr>
<td>20</td>
<td>0.506</td>
<td>0.854</td>
</tr>
</tbody>
</table>

DCF+SCFs, Depth 5+10, 4.5MB, 85%
SCFs, Depth 15, 110MB, 50%
93% per-frame accuracy
Leave One Subject Out

<table>
<thead>
<tr>
<th></th>
<th>Pinch Open</th>
<th>Pinch Close</th>
<th>Pointing</th>
<th>Gun</th>
<th>Splayed Hand</th>
<th>Flat Hand</th>
<th>No-Gesture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pinch Open</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pinch Close</strong></td>
<td>0.88</td>
<td>0.03</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.02</td>
</tr>
<tr>
<td><strong>Pointing</strong></td>
<td>0.0</td>
<td>0.93</td>
<td>0.05</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Gun</strong></td>
<td>0.04</td>
<td>0.0</td>
<td>0.9</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>Splayed Hand</strong></td>
<td>0.02</td>
<td>0.02</td>
<td>0.95</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>Flat Hand</strong></td>
<td>0.01</td>
<td>0.99</td>
<td>0.01</td>
<td>0.99</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>No-Gesture</strong></td>
<td>0.05</td>
<td>0.03</td>
<td>0.03</td>
<td>0.0</td>
<td>0.0</td>
<td>0.01</td>
<td>0.85</td>
</tr>
</tbody>
</table>
Applications
Not Only Mobile Phones
Not Only Mobile Phones
Limitations

• Low-lighting still a problem

• Having more robust segmentation would help

• Gesture with non-unique contour are still a problem

• Mostly static gestures
Contributions

• Robust and real-time gesture recognition on unmodified mobile devices

• Extending multi-layered Random Forest framework for low-memory devices

• Variety of different mobile platforms, including phones, tablets and even smart-watches

• Variety of interaction techniques and scenarios where in-air gestures complement touch
Live Demo
Thank You!

In-air Gestures
Around Unmodified Mobile Devices

Jie Song¹, Gabor Soros¹, Fabrizio Pece*,
Sean Fanello², Shahram Izadi²,
Cem Keskin², Otmar Hilliges¹

¹ETH Zurich ²Microsoft Research

{jsong|gabor.soros|fabrizio.pece|otmar.hilliges}@inf.ethz.ch
{seanfa|shahrami|cemke}@microsoft.com