HAND GESTURE RECOGNITION WITH DEPTH DATA

Fabio Dominio, Mauro Donadeo, Giulio Marin, Pietro Zanuttigh and Guido M. Cortelazzo
Hand gesture recognition with depth data

Hand gesture recognition is an intriguing problem with many applications.

Large amount of research on hand gesture recognition from image and videos but it remains a challenging task due to the complex geometry of the hand and to the inter-occlusions.

Now Depth data is easily available from low cost devices.

Depth data offers a very accurate representation of the hand shape and allows to improve the accuracy of gesture recognition schemes.
Overview of the Approach

- **Depth data**
  - Extraction of the hand region
  - Hand samples
    - Color data
    - PCA
    - Circle fitting on the palm
    - Extraction of hand parts
      - Finger samples
      - Palm samples
      - Wrist samples (discarded)

- **Feature extraction**
  - Distance features extraction
  - Curvature features extraction

- **SVM classification**
  - Training data
  - SVM
  - Recognized gesture
Both color and depth data are used for hand recognition.

Start from closest point $X_{\text{min}}$ (if it is an isolated point a new point is selected).

Thresholding for initial hand estimation:
- On the **depth** value
- On the **distance** from the closest point in 3D space

Mathematically:

$$\mathcal{H} = \{ X_i | D(X_i) < D(X_{\text{min}}) + T_h \wedge \| X_i - X_{\text{min}} \| < T_{h2} \}$$
Extraction of the Hand (2)

- Hand compatibility check:
  - Detected object size must be compatible with the hand
  - Detected object color in the CIELAB space must be compatible with the skin color
  - Wrist and part of the forearm could be included

- Initial palm center position detection
  - Low pass Gaussian filtering of the mask to find highest density region
  - $\sigma$ depends on the hand distance
  - Center of the palm: point of highest density closest to $X_{min}$
Extraction of the Palm and Fingers

- A circle is fitted on the palm starting from the estimated palm center
- Search for the maximum size circle that can be fitted on the palm area
  - 95% of the circle must be inside the detected region
- Refinement of the palm center position
  - 2 iterated phases: move/enlarge
- Subdivision into palm, fingers and wrist regions
- Improvement of the palm recognition with ellipse fitting [1]

The rough orientation of the hand is detected using Principal Component Analysis (PCA).

A plane is fitted on the hand’s palm using SVD and RANSAC:

- RANSAC ensures robustness to Kinect’s artifacts.

A new reference system is built.
Two different types of features are extracted:

1. **Distance features** computed on the finger samples
2. **Curvature features** computed on the hand contour

Two additional types of features (**palm area features and elevations from the hand’s plane**) have been added in journal extension on Pattern Recognition Letters [2]

We consider the distances of the finger samples from the hand centroid *in the 3D space* for each angular direction.

An histogram with the maximum value for each direction is built.

Alignment with reference templates for precise hand orientation.

Histograms flipping to handle left/right hand and palm/dorsum facing the camera.

\[
L(\theta_q) = \max_{\theta_q - \frac{\Delta}{2} < \theta_{x_i} \leq \theta_q + \frac{\Delta}{2}} d_{x_i}
\]

\[
\Delta_g = \arg \max_{\Delta} (\rho(L(\theta), L_g^*(\theta + \Delta)))
\]

\[
\Delta_{g_{rev}} = \arg \max_{\Delta} (\rho(L(-\theta), L_g^*(\theta + \Delta)))
\]
Angular directions are divided into regions corresponding to the fingers of interest in the considered gesture.

Feature values are the normalized maxima in the region corresponding to each finger.

There is one feature for each finger in each gesture hypotheses (i.e., there can be up to G*5 features).
### Distance Features: Examples

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Rep. 1</th>
<th>Rep. 2</th>
<th>Rep. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Gesture Image" /></td>
<td><img src="graph1.png" alt="Graph" /></td>
<td><img src="graph2.png" alt="Graph" /></td>
<td><img src="graph3.png" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image2.png" alt="Gesture Image" /></td>
<td><img src="graph4.png" alt="Graph" /></td>
<td><img src="graph5.png" alt="Graph" /></td>
<td><img src="graph6.png" alt="Graph" /></td>
</tr>
<tr>
<td><img src="image3.png" alt="Gesture Image" /></td>
<td><img src="graph7.png" alt="Graph" /></td>
<td><img src="graph8.png" alt="Graph" /></td>
<td><img src="graph9.png" alt="Graph" /></td>
</tr>
</tbody>
</table>
Curvature Features (1)

- Computed on the edges of the hand region (palm and fingers)
- Multi-scale descriptor of the curvature of the edges
- Curvature $V(x_i, s)$: ratio between hand area inside a circular mask centered on $x_i$ and the mask size
  - $V < 0.5$: convex region
  - $V = 0.5$: straight edge
  - $V > 0.5$: concave region

$$V(X_i, s) = \frac{\sum_{x \in M_s(x_i)} h_m(X)}{|M_s(X_i)|}$$
Curvature Features (2)

- Feature value: number of samples with a certain curvature at the selected scale level
- Curvature values interval divided into B bins of equal size
- Feature vector: 2D array with count of samples with a certain curvature at a certain scale level

\[ n_{b,s} = \{ X_i \mid \frac{(b - 1)}{B} < V(X_i, s) \leq \frac{b}{B} \} \]
## Curvature Features: Examples

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Rep. 1</th>
<th>Rep. 2</th>
<th>Rep. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="img1" alt="Gesture Image" /></td>
<td><img src="img2" alt="Rep. 1 Image" /></td>
<td><img src="img3" alt="Rep. 2 Image" /></td>
<td><img src="img4" alt="Rep. 3 Image" /></td>
</tr>
<tr>
<td><img src="img5" alt="Gesture Image" /></td>
<td><img src="img6" alt="Rep. 1 Image" /></td>
<td><img src="img7" alt="Rep. 2 Image" /></td>
<td><img src="img8" alt="Rep. 3 Image" /></td>
</tr>
<tr>
<td><img src="img9" alt="Gesture Image" /></td>
<td><img src="img10" alt="Rep. 1 Image" /></td>
<td><img src="img11" alt="Rep. 2 Image" /></td>
<td><img src="img12" alt="Rep. 3 Image" /></td>
</tr>
</tbody>
</table>
Feature vectors are built by concatenating the different distance and curvature features.
- Distance features: one for each relevant finger in each gesture hypothesis.
- Curvature features: one for each curvature bin and scale level.
- Classification with a multi-class support vector machine:
  - One-against-one approach
  - Kernel: Gaussian Radial Basis Function (RBF)
  - Grid-search with cross validation for parameters tuning
**Experimental Results**

<table>
<thead>
<tr>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>G8</th>
<th>G9</th>
<th>G10</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="Gesture Image" /></td>
<td><img src="image2" alt="Gesture Image" /></td>
<td><img src="image3" alt="Gesture Image" /></td>
<td><img src="image4" alt="Gesture Image" /></td>
<td><img src="image5" alt="Gesture Image" /></td>
<td><img src="image6" alt="Gesture Image" /></td>
<td><img src="image7" alt="Gesture Image" /></td>
<td><img src="image8" alt="Gesture Image" /></td>
<td><img src="image9" alt="Gesture Image" /></td>
<td><img src="image10" alt="Gesture Image" /></td>
</tr>
<tr>
<td><img src="image11" alt="Graph Image" /></td>
<td><img src="image12" alt="Graph Image" /></td>
<td><img src="image13" alt="Graph Image" /></td>
<td><img src="image14" alt="Graph Image" /></td>
<td><img src="image15" alt="Graph Image" /></td>
<td><img src="image16" alt="Graph Image" /></td>
<td><img src="image17" alt="Graph Image" /></td>
<td><img src="image18" alt="Graph Image" /></td>
<td><img src="image19" alt="Graph Image" /></td>
<td><img src="image20" alt="Graph Image" /></td>
</tr>
<tr>
<td><img src="image21" alt="Heatmap Image" /></td>
<td><img src="image22" alt="Heatmap Image" /></td>
<td><img src="image23" alt="Heatmap Image" /></td>
<td><img src="image24" alt="Heatmap Image" /></td>
<td><img src="image25" alt="Heatmap Image" /></td>
<td><img src="image26" alt="Heatmap Image" /></td>
<td><img src="image27" alt="Heatmap Image" /></td>
<td><img src="image28" alt="Heatmap Image" /></td>
<td><img src="image29" alt="Heatmap Image" /></td>
<td><img src="image30" alt="Heatmap Image" /></td>
</tr>
</tbody>
</table>

- Gesture database from the work of Ren et Al [3]
- 1000 samples: 10 gestures, 10 people and 10 repetitions for each gesture
- Results on a second more challenging dataset are included in the journal extension [2]


The database has been divided into 800 samples for training and 200 for testing.

2 Subdivision modalities
1. Random subdivision (user training)
2. 8 people for testing and 2 for training (generic training)

- Grid search with cross validation for optimal parameters extraction
**Experimental Results**

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Training with users</td>
</tr>
<tr>
<td>Distance features</td>
<td>96%</td>
</tr>
<tr>
<td>Curvature features</td>
<td>97,5%</td>
</tr>
<tr>
<td>Distance + Curvature</td>
<td><strong>99,5%</strong></td>
</tr>
<tr>
<td>Shape context [1]</td>
<td>83,2%</td>
</tr>
<tr>
<td>Near-convex Dec.+FEMD [3]</td>
<td>90,6%</td>
</tr>
<tr>
<td>Thresholding Dec.+FEMD [3]</td>
<td>93,9%</td>
</tr>
</tbody>
</table>

- Combined use of the two features: better performances
- 99,5% accuracy with user training
- Generic training more challenging, but the combined use of the two features leads to very good results (98,5%)
- Large performance improvement w.r.t. [3]

## Confusion Matrices

### Distance

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>G8</th>
<th>G9</th>
<th>G10</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G3</td>
<td>0</td>
<td>0.05</td>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0.05</td>
<td>0.15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G7</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.85</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>G8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>G9</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.85</td>
</tr>
<tr>
<td>G10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.85</td>
</tr>
</tbody>
</table>

### Curvature

<table>
<thead>
<tr>
<th></th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th>G5</th>
<th>G6</th>
<th>G7</th>
<th>G8</th>
<th>G9</th>
<th>G10</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G3</td>
<td>0</td>
<td>0</td>
<td>0.8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.15</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.95</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G5</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0.95</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G7</td>
<td>0</td>
<td>0</td>
<td>0.25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.95</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>G8</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0.7</td>
<td>0</td>
</tr>
<tr>
<td>G9</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>G10</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.95</td>
</tr>
</tbody>
</table>

### Combined
• More challenging dataset (from extended version of the work [2])
• 1680 samples: 12 gestures, 14 people and 10 repetitions for each gesture
• Accuracy of 95% with user training and 89.6% with generic training
• Up to 97.6% and 93.8% with 4 features

Conclusions

- The hand is reliably extracted from the color and depth data
- Recognition of the palm, fingers and hand orientation
- Reliable feature descriptors based on 3D measures
- Distance and curvature features capture different clues: they are complementary
- Real-time computation (10 fps)
- Very high accuracy on datasets from the literature
Future Research

- Additional feature descriptors from depth and color data
  - Elevation features
  - Palm area features
  - Color-based features

- Better palm area identification
- Recognition of multiple interacting hands
- Advanced machine learning strategies
- Extension to dynamic gestures recognition
Thanks for your attention

For datasets and further information on our research:
visit our website http://lttm.dei.unipd.it