Real-time gesture recognition from depth data through key poses learning and decision forests

Leandro Miranda
Thales Vieira (presenter)
Dimas Martinez
Mathematics, UFAL

Thomas Lewiner
Mathematics, PUC-Rio

Antonio W. Vieira
Mario F. M. Campos
Computer Science, UFMG
Human Gesture Recognition
Human Gesture Recognition

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Human Gesture Recognition

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Human Gesture Recognition

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Current Scenario

Popularization of real time depth sensors

Development of high quality Natural User Interfaces (NUI)

Miranda et al., 2012
Current Scenario

Popularization of real time depth sensors

Challenging task! Gestures performed at different speeds and/or sequence of poses

Miranda et al., 2012
Our approach: key poses learning

Gestures can be characterized by a few extreme poses!

✓ Real-time gesture learning and recognition
✓ Ideal for the average inexperienced user
Outline

1. Related Work
2. Overview
3. Joint-angles Representation
4. Key Pose Statistical Learning
5. Gesture Recognition Through Decision Forests
6. Results
Related Work

Local methods

- Li et al. (2010)

Global methods

- Lv and Nevatia (2007)

Parametric methods

- Raptis et al. (2011)

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
**Overview**

Pose descriptor extraction

(key pose learning machine)

Gesture learning machine

(Decision forest)

Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Overview: training key poses

pose descriptor extraction

key pose learning machine

kinect \( \rightarrow \) \((x_1, \cdots, x_{15})\) \(\rightarrow\) \((\theta_1, \varphi_1, \cdots, \theta_9, \varphi_9, \eta)\)
Overview: recognizing key poses

pose descriptor extraction

kinect $\to (x_1, \cdots, x_{15}) \to (\theta_1, \varphi_1, \cdots, \theta_9, \varphi_9, \eta)$

key pose learning machine

training set
multi-class SVM

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Overview: recognizing key poses

**Pose descriptor extraction**

1. Kinect
2. \((x_1, \ldots, x_{15})\)
3. \((\theta_1, \varphi_1, \ldots, \theta_9, \varphi_9, \eta)\)

**Key pose learning machine**

- Training set
- Multi-class SVM

**Key pose**

*Miranda et al., 2012*
Overview: training gestures

Pose descriptor extraction

key pose learning machine

kinect $\rightarrow (x_1, \cdots, x_{15}) \rightarrow (\theta_1, \varphi_1, \cdots, \theta_9, \varphi_9, \eta)$

training set

multi-class SVM
Overview: training gestures

Pose descriptor extraction

Kinect $\rightarrow (x_1, \ldots, x_{15}) \rightarrow (\theta_1, \varphi_1, \ldots, \theta_9, \varphi_9, \eta)$

Key pose learning machine

Training set $\rightarrow$ Multi-class SVM

Key pose
Overview: training gestures

pose descriptor extraction

(\(x_1, \cdots, x_{15}\))

(\(\theta_1, \varphi_1, \cdots, \theta_9, \varphi_9, \eta\))

gesture learning machine

decision forest

Key pose learning machine

training set

multi-class SVM

key pose
Overview: recognizing gestures

pose descriptor extraction

kinect \( \rightarrow \) \((x_1, \cdots, x_{15})\) \(\rightarrow\) \((\theta_1, \varphi_1, \cdots, \theta_9, \varphi_9, \eta)\)

key pose learning machine

training set \(\downarrow\) multi-class SVM \(\downarrow\) key pose

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Overview: recognizing gestures

Pose descriptor extraction

kinect \rightarrow (x_1, \ldots, x_{15}) \rightarrow (\theta_1, \varphi_1, \ldots, \theta_9, \varphi_9, \eta)

Gesture learning machine

key pose learning machine

training set

multi-class SVM

key pose

decision forest

training set

buffer

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Overview: recognizing gestures

pose descriptor extraction

kinect → \((x_1, \cdots, x_{15})\) → \((\theta_1, \varphi_1, \cdots, \theta_9, \varphi_9, \eta)\)

gesture learning machine

decision forest

gesture → \(k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10}\) → \(g_1, g_2, g_3, g_4, g_5, g_6\)

training set

key pose learning machine

training set → multi-class SVM

key pose

\((k_1, t_1) \rightarrow (k_2, t_2) \rightarrow \cdots \rightarrow (k_n, t_n)\)

buffer
Overview

Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012

Overview

pose descriptor extraction

kinect

\( (x_1, \cdots, x_{15}) \)

\( (\theta_1, \varphi_1, \cdots, \theta_9, \varphi_9, \eta) \)

gesture learning machine

decision forest

training set

multi-class SVM

key pose

buffer
Skelettons from Kinect Sensor

Real-time depth sensing system streaming depth data and skeletons at 30fps

Miranda et al., 2012 

Real-time gesture recognition from depth data through key poses learning and decision forests
Joint-Angles Pose Descriptor

**Objective:** Concise and invariant representation of relevant pose information.

Improvement of Raptis et al (2011) local spherical coordinates.

$$(x_1, x_2, \ldots, x_{15}) \in \mathbb{R}^{45}$$

$$(\theta_1, \varphi_1, \ldots, \theta_9, \varphi_9, \eta) \in \mathbb{R}^{19}$$

1st degree joints: elbows, knees and head
2nd degree joints: hands, feet.
How to compute the local bases?

1st degree joints:
How to compute the local bases?

1st degree joints:
How to compute the local bases?

1st degree joints:

θ - angle between $\vec{u}$ and $\vec{w}$

φ - angle between $\vec{t}$ and the projection of $\vec{w}$ in $\pi$
How to compute the local bases?

2nd degree joints:

Rotate $\{\vec{u}, \vec{r}, \vec{t}\}$ by

$$\beta = \arccos(\vec{w}, \vec{u})$$

around

$$b = \vec{w} \times \vec{u}$$
How to compute the local bases?

2nd degree joints:

\[ \beta = \arccos(\vec{w}, \vec{u}) \]

Rotate \( \{ \vec{u}, \vec{r}, \vec{t} \} \) by around

\[ b = \vec{w} \times \vec{u} \]

Real-time gesture recognition from depth data through key poses learning and decision forests
How to compute the local bases?

2nd degree joints:

Rotate \( \{ \vec{u}, \vec{r}, \vec{t} \} \) by

\[
\beta = \arccos(\vec{w}, \vec{u})
\]

around

\[
b = \vec{w} \times \vec{u}
\]

\( \theta \) - angle between rotated \( \vec{u} \) and \( \vec{q} \)
\( \varphi \) - angle between rotated \( \vec{t} \) and the projection of \( \vec{q} \) in \( \pi \)
Real-time gesture recognition from depth data through key poses learning and decision forests.

Overview

pose descriptor extraction

kinect \( \rightarrow \) \((x_1, \ldots, x_{15})\)

\((\theta_1, \varphi_1, \ldots, \theta_9, \varphi_9, \eta)\)

pose descriptor extraction

gesture learning machine

decision forest

\((k_1, t_1) \rightarrow (k_2, t_2) \rightarrow \cdots \rightarrow (k_n, t_n)\)

key pose learning machine

training set

multi-class SVM

key pose

buffer

Miranda et al., 2012
Overview

Real-time gesture recognition from depth data through key poses learning and decision forests
Supervised Learning Machine

Predefined key pose classes: \( \mathcal{K} = \{ k_1, k_2, \ldots, k_{|\mathcal{K}|} \} \)
Supervised Learning Machine

Predefined key pose classes: $\mathcal{K} = \{ k_1, k_2, \ldots, k_{|\mathcal{K}|} \}$

- $(v_1, c_1)$
- $(v_2, c_2)$
- \[ \ldots \]
- $(v_n, c_n)$

Training Set

$(\theta_1, \varphi_1, \ldots, \theta_9, \varphi_9, \eta) \in \mathbb{R}^{19}$, $c_1 \in \mathcal{K}$

Machine
Supervised Learning Machine

Predefined key pose classes: $\mathcal{K} = \{k_1, k_2, \ldots, k_{|\mathcal{K}|}\}$

Training Set

$\{(v_1, c_1), (v_2, c_2), \ldots, (v_n, c_n)\}$

$(\theta_1, \varphi_1, \ldots, \theta_9, \varphi_9, \eta) \in \mathbb{R}^{19}, \quad c_1 \in \mathcal{K}$

Machine

$k_i$
Support Vector Machines (SVM)

Binary classifier

\[ \hat{g} : \mathbb{R}^k \rightarrow \{-1, 1\} \]
\[ v \rightarrow \text{sign}\left(\hat{f}(v)\right) = \{-1, 1\} \]

\[ \hat{f}(v) = \sum_j^\beta \alpha_j s_j \langle \varphi(v_j), \varphi(v) \rangle + b \]

\[
\begin{array}{l}
\text{MAX}_{w,\gamma} \quad \gamma - C \sum_{i=1}^l \varepsilon_i \\
\text{subject to} \quad y_i \langle w, \Phi(x_i) \rangle \geq \gamma - \varepsilon_i, \varepsilon_i \geq 0, \quad \|w\|^2 = \\
\end{array}
\]

✓ Non-linear classification
✓ Efficiently computed for small training sets
Multi-class SVM formulation

One-versus-all approach

One binary classifier for each key pose $\mathbf{p} \in \mathcal{K}$:

$$\hat{f}_\mathbf{p}(\mathbf{v}) = \sum_{j \in \text{SV}} \alpha_j \psi_\mathbf{p}(c_j) \phi(\mathbf{v}_j, \mathbf{v}) + b,$$

where

$$\psi_\mathbf{p}(c) = \begin{cases} 1 & \text{if } c = \mathbf{p}, \\ -1 & \text{otherwise}, \end{cases}$$

$$\phi(\mathbf{v}_1, \mathbf{v}_2) = \exp \left( -\frac{\|\mathbf{v}_2 - \mathbf{v}_1\|^2}{2\sigma^2} \right)$$

Voting process:

$$\hat{f}(\mathbf{v}) = \begin{cases} \mathbf{q} = \arg \max_\mathbf{p} \hat{f}_\mathbf{p}(\mathbf{v}) & \text{if } \hat{f}_\mathbf{q}(\mathbf{v}) > 0, \\ -1 & \text{otherwise}. \end{cases}$$
Overview

Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Overview

Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Gestures as key pose sequences

Gesture representation: \( g = \{k_1, k_2, \cdots, k_{n_g}\}, \quad k_i \in \mathcal{K}. \)
Gestures as key pose sequences

Gesture representation: \( g = \{ k_1, k_2, \cdots, k_{n_g} \}, \quad k_i \in \mathcal{K}. \)

Training session:
Gestures as key pose sequences

Gesture representation: \( g = \{k_1, k_2, \cdots, k_{n_g}\}, \quad k_i \in \mathcal{K}. \)

Training session:

Miranda et al., 2012
Decision Forests

Each node represents a key pose

One tree per key pose

Each root-leaf path represents a gesture stored back-to-front

Two paths may represent the same gesture

Miranda et al., 2012
Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Real-time gesture recognition

key pose learning machine

\[ k_{i+n} \]

buffer

\[ k_i, k_{i+1}, \ldots, k_{i+n} \]

decision forest

\[ (k_1, t_1), (k_2, t_2), \ldots, (k_n, t_n) \]

Miranda et al., 2012
Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Real-time gesture recognition: Example

Miranda et al., 2012
Real-time gesture recognition: Example

Real-time gesture recognition from depth data through key poses learning and decision forests

Miranda et al., 2012
Real-time gesture recognition: Example

Miranda et al., 2012
Real-time gesture recognition: Example

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Real-time gesture recognition: Example

Miranda et al., 2012
Real-time gesture recognition: Example

$$(k_1, t_1) \rightarrow (k_2, t_2) \rightarrow \ldots \rightarrow (k_n, t_n)$$

Miranda et al., 2012

Real-time gesture recognition from depth data through key poses learning and decision forests
Time constraints

Time vector: interval \( t = [t_1, t_2, \cdots, t_{n-1}] \) between consecutive key poses

Time test

for each time vector \( t_i \) found on the leaf

\[
\text{if } \| t_i - t \|_\infty > T \\
\text{discard } t_i
\]

return \( g_i \) that minimizes \( \| t_i - t \|_1 \)
Results
Experiments Setup

One trainer

18 trained key poses (approx. 30 examples per key pose)

10 trained gestures (approx. 10 executions per gesture)

Miranda et al., 2012
# Key pose recognition: robustness

10 inexperienced individuals performed trained key poses 10 times

<table>
<thead>
<tr>
<th>key pose</th>
<th>id</th>
<th>$u_1$</th>
<th>$u_2$</th>
<th>$u_3$</th>
<th>$u_4$</th>
<th>$u_5$</th>
<th>$u_6$</th>
<th>$u_7$</th>
<th>$u_8$</th>
<th>$u_9$</th>
<th>$u_{10}$</th>
<th>$u_{10}'$</th>
<th>total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>$k_1$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>100.00</td>
</tr>
<tr>
<td>Right Hand Right</td>
<td>$k_2$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>98.18</td>
</tr>
<tr>
<td>Left Hand Left</td>
<td>$k_3$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>98.18</td>
</tr>
<tr>
<td>Arms Open</td>
<td>$k_4$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>7</td>
<td>10</td>
<td>93.63</td>
</tr>
<tr>
<td>Right Hand Front</td>
<td>$k_5$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>10</td>
<td></td>
<td>95.45</td>
</tr>
<tr>
<td>Left Hand Front</td>
<td>$k_6$</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td>99.09</td>
</tr>
<tr>
<td>Both Hands Front</td>
<td>$k_7$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Right Hand Up</td>
<td>$k_8$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Left Hand Up</td>
<td>$k_9$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Both Hands Up</td>
<td>$k_{10}$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Right Hand 90°</td>
<td>$k_{11}$</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td></td>
<td>95.45</td>
</tr>
<tr>
<td>Left Hand 90°</td>
<td>$k_{12}$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td>91.81</td>
</tr>
<tr>
<td>Both Hands 90°</td>
<td>$k_{13}$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td>100.00</td>
</tr>
<tr>
<td>Inclined Front</td>
<td>$k_{14}$</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td>89.09</td>
</tr>
<tr>
<td>Hands-on-Hip Crossed</td>
<td>$k_{15}$</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td></td>
<td>84.54</td>
</tr>
<tr>
<td>Hand-On-Hip</td>
<td>$k_{16}$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td></td>
<td>99.09</td>
</tr>
<tr>
<td>Hands on Head</td>
<td>$k_{17}$</td>
<td>9</td>
<td>10</td>
<td>10</td>
<td>8</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td></td>
<td>90.00</td>
</tr>
<tr>
<td>Right Hand 90° Back</td>
<td>$k_{18}$</td>
<td>8</td>
<td>10</td>
<td>9</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>8</td>
<td>77.27</td>
</tr>
</tbody>
</table>

**Average recognition rate:** 94.84%

Miranda et al., 2012
Key pose recognition: stability

Out-of-sample tests:

1. Remove 20% of training set data;
2. Compute SVM classifier;
3. Try to classify removed training data.

Results after 10 experiments:

False classifications: 4.16%
Unclassified key poses: 3.45%
Key pose recognition
# Gesture recognition

10 inexperienced individuals performed trained gestures 10 times

<table>
<thead>
<tr>
<th>gesture</th>
<th>id</th>
<th>key pose seq.</th>
<th>rec. rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-Clap</td>
<td>$g_1$</td>
<td>$k_1, k_4, k_7$</td>
<td>99%</td>
</tr>
<tr>
<td>Open Arms</td>
<td>$g_2$</td>
<td>$k_1, k_7, k_4$</td>
<td>96%</td>
</tr>
<tr>
<td>Turn Next Page</td>
<td>$g_3$</td>
<td>$k_1, k_2, k_5, k_1$</td>
<td>83%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_1, k_6, k_3, k_1$</td>
<td></td>
</tr>
<tr>
<td>Turn Previous Page</td>
<td>$g_4$</td>
<td>$k_1, k_5, k_2, k_1$</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_1, k_3, k_6, k_1$</td>
<td></td>
</tr>
<tr>
<td>Raise Right Arm Laterally</td>
<td>$g_5$</td>
<td>$k_1, k_2, k_8$</td>
<td>80%</td>
</tr>
<tr>
<td>Lower Right Arm Laterally</td>
<td>$g_6$</td>
<td>$k_8, k_2, k_1$</td>
<td>78%</td>
</tr>
<tr>
<td>Good Bye</td>
<td>$g_7$</td>
<td>$k_1, k_{11}$</td>
<td>92%</td>
</tr>
<tr>
<td>(time constraint: 1 sec.)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese Greeting</td>
<td>$g_8$</td>
<td>$k_1, k_{14}, k_1$</td>
<td>100%</td>
</tr>
<tr>
<td>Put Hands Up Front</td>
<td>$g_9$</td>
<td>$k_1, k_5, k_18$</td>
<td>96%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_1, k_5, k_8$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_1, k_5, k_{11}, k_8$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$k_1, k_8$</td>
<td></td>
</tr>
<tr>
<td>Put Hands Up Laterally</td>
<td>$g_{10}$</td>
<td>$k_1, k_4, k_{10}$</td>
<td>100%</td>
</tr>
</tbody>
</table>
Gesture recognition
Performance

Preprocessing bottleneck: computing SVM classifiers

For a training set of 2,000 key pose examples of 18 classes:
18 functions were computed in 3.9 secs

Negligible performance during training/recognition phases

Usually very low tree depths
Comparison

Dataset from Li et al (2010): 20 gestures, 10 individuals, 3 executions

<table>
<thead>
<tr>
<th>AS1</th>
<th>AS2</th>
<th>AS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal arm wave</td>
<td>High arm wave</td>
<td>High throw</td>
</tr>
<tr>
<td>Hammer</td>
<td>Hand catch</td>
<td>Forward kick</td>
</tr>
<tr>
<td>Forward punch</td>
<td>Draw x</td>
<td>Side kick</td>
</tr>
<tr>
<td>High throw</td>
<td>Draw tick</td>
<td>Jogging</td>
</tr>
<tr>
<td>Hand clap</td>
<td>Draw circle</td>
<td>Tennis swing</td>
</tr>
<tr>
<td>Bend</td>
<td>Two hand wave</td>
<td>Tennis serve</td>
</tr>
<tr>
<td>Pickup &amp; throw</td>
<td>Side boxing</td>
<td>Pickup &amp; throw</td>
</tr>
</tbody>
</table>

Cross-subject test:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AS1</td>
<td>72.9%</td>
<td>84.7%</td>
<td>93.5%</td>
</tr>
<tr>
<td>AS2</td>
<td>71.9%</td>
<td>81.3%</td>
<td>52.0%</td>
</tr>
<tr>
<td>AS3</td>
<td>79.2%</td>
<td>88.4%</td>
<td>95.4%</td>
</tr>
<tr>
<td>Average</td>
<td>74.7%</td>
<td>84.8%</td>
<td>80.3%</td>
</tr>
</tbody>
</table>
Comparison

Dataset from Li et al (2010): 20 gestures, 10 individuals, 3 executions

<table>
<thead>
<tr>
<th>AS1</th>
<th>AS2</th>
<th>AS3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal arm wave</td>
<td>High arm wave</td>
<td>High throw</td>
</tr>
<tr>
<td>Hammer</td>
<td>Hand catch</td>
<td>Forward kick</td>
</tr>
<tr>
<td>Forward punch</td>
<td>Draw x</td>
<td>Side kick</td>
</tr>
<tr>
<td>High throw</td>
<td>Draw tick</td>
<td>Jogging</td>
</tr>
<tr>
<td>Hand clap</td>
<td>Draw circle</td>
<td>Tennis swing</td>
</tr>
<tr>
<td>Bend</td>
<td>Two hand wave</td>
<td>Tennis serve</td>
</tr>
<tr>
<td>Pickup &amp; throw</td>
<td>Side boxing</td>
<td>Pickup &amp; throw</td>
</tr>
</tbody>
</table>

**Cross-subject test:**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>AS1</td>
<td>72.9%</td>
<td>84.7%</td>
<td>93.5%</td>
</tr>
<tr>
<td>AS2</td>
<td>71.9%</td>
<td>81.3%</td>
<td>52.0%</td>
</tr>
<tr>
<td>AS3</td>
<td>79.2%</td>
<td>88.4%</td>
<td>95.4%</td>
</tr>
<tr>
<td>Average</td>
<td>74.7%</td>
<td>84.8%</td>
<td>80.3%</td>
</tr>
</tbody>
</table>

Delicate gestures
Limitations

- Robustness issues
  - Skeleton tracking
  - Delicate gestures

- Key pose design not the friendliest solution
Future Work

✓ Automatic key pose generation

✓ Work on skeleton tracking algorithms (More than 1 Kinect?)

✓ Improve time constrained gesture recognition

✓ Take into account key pose descriptor periodicity
Thank you for your attention!
Thank you for your attention!
Thank you for your attention!

Questions?