Real-time Arm Skeleton Tracking and Gesture Inference Tolerant to Missing Wearable Sensors

Yang Liu¹, Zhenjiang Li¹, Zhidan Liu², Kaishun Wu²
City University of Hong Kong¹, Shenzhen University²
Understanding Human Arm Motions

- How is the arm moving? 
  Skeleton tracking

- What is the meaning of this arm motion? 
  Motion inference

Running
Elderly Care

**Elderly diseases**
- Parkinson
- Alzheimer

Problems with arm
- Slow motion
- Repeated motion
- Instability
  - ...

Last treatment

Several weeks

Next treatment

Unified Parkinson’s Disease Rating Scale

Hand Movements (Patient opens and closes hands in rapid succession)

- 0 = Normal.
- 1 = Mild slowing and/or reduction in amplitude.
- 2 = Moderately impaired; slowness and effort fatigue. May have movement.
- 3 = Severely impaired. From initiating movements to movement.
- 4 = Can barely perform the action.
Other Applications

Template → Compare → User’s skeleton

α → α'
Other Applications

80USD/hour

Template

User’s skeleton

α

α’

Compare

HCl

Smart home

Smart car

Gaming
Existing Solutions

- Service coverage
- System cost
- Privacy

- Convenience
- User-friendly
Existing Solutions

- Service coverage
- System cost
- Privacy
- Convenience
- User-friendly

Few
Existing Solutions

- Service coverage
- System cost
- Privacy
- Convenience
- User-friendly

Few
Key Problem

Elbow

On body

(Fixed offset)

Wrist
Key Problem

How?

Elbow
Tracking Principle

For a given **wrist orientation**, possible elbow locations are **within a limited range** [1].

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.
Tracking Principle

Ranges across time [1]

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.
Latency

Time delay of existing work [1]

- 10s: 98.2s
- 30s: 289.3s
- 1min: 9.1min

10x on desktop

Activity duration
Recovery delay

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.
Latency

- Our solution [ArmTroi]:
  - HMM state reconstruction
  - Hierarchical search

- Real-time
- Without impairing accuracy

Time delay of existing work [1]

10x on desktop

Activity duration
Recovery delay

Time

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.
Our idea: exclude the unlikely locations using as little effort as possible
Focus on more likely candidates
Hierarchical Search

- **First-layer search**
- **Second-layer search**

- **Tracked Elbow Location**
- **Tracked Center Point**

**Time complexity:**

\[
O \left( \left( \frac{N}{n_1} \right)^2 T + (n_1)^2 T \right)
\]

- **Real-time**
- **Without impairing accuracy**

**Original size**

**Size of region**
Understanding Human Arm Motions

• How is the arm moving?
  Skeleton tracking

• What is the meaning of this arm motion?
  Motion inference

Running
Motion Inference

6 combinations of missing inputs

Non-scalable

Cost-inefficient
Motion Inference

6 combinations of missing inputs

Handle all combinations using one network?

Non-scalable

Cost-inefficient
Our idea

- Adaptive design

\[ a_{11} + a_{12} + a_{13} = 1 \]
Attention-based network adaption

Features: $z^t = \{z_1^t, z_2^t, z_3^t\}$

Input: $x^t$

$z^t \rightarrow x^t$: Weighted fusion

$x^t = \phi(\{z_r^t\}, \{\alpha_r^t\}), r = 1, 2, 3$

Updated weights

- Weight update
  - aligning with the activity descriptor $h_{t-1}$

$\alpha^t = f_{att}(z^t, h^{t-1})$

$\text{att}^t = \text{Relu}(z^t + Wh^{t-1} + b)$

$\alpha^t = f_{\text{soft-max}}(U\text{att}^t)$
ArmTroi Implementation

**Skeleton Tracking**
- Kinetic Model
- Point Clouds
- Raw Data
- Skeleton Recover
  - Acceleration
  - Arm
  - Torso
- Skeletons

**Gesture Inference**
- DNN
- Network Structure Design
- Attention-based Adaptation

**Applications**
- Elderly Care
- E-Health
- HCI
- Behavior Analysis

Elderly Care Applications
Experiment setup

- Participants: 7 volunteers
- Dataset:

<table>
<thead>
<tr>
<th>Categories</th>
<th>Gestures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daily gestures</td>
<td>shake hands, make a call, open a door, drink water</td>
</tr>
<tr>
<td>Free-weight</td>
<td>front raise (a/p), biceps curl (a/p), bent over single arm, chest fly (i/s), bench press (i/s), lateral raise</td>
</tr>
<tr>
<td>Customized</td>
<td>push, pull, circle</td>
</tr>
</tbody>
</table>

Table 1: Targeted gestures in ArmTroi. The a, p, i, s stand for alternating, in parallel, incline and sitting, respectively.

- Training: Intel i7-6700 CPU and Nvidia GTX 1080Ti GPU
- Running: SAMSUNG Galaxy S7
Evaluation

- Skeleton tracking

- ArmTrak [1]
  - Elbow: 12.94cm
  - Wrist: 14.91cm

- ArmTroi
  - Elbow: 10.53cm
  - Wrist: 12.94cm

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.
Evaluation

- Skeleton tracking

- ArmTrak [1]
  - Elbow: 12.94cm
  - Wrist: 14.91cm

- ArmTroi
  - Elbow: 10.53cm
  - Wrist: 12.94cm

- Our latency
  - Desktop: 0.15x
  - Phone: 0.47x

[1] “I am a smartwatch and I can track my user’s arm”, in Proc. of ACM MobiSys, 2016.
Evaluation

• Motion inference

- Baseline: MULT
  - Each combination of missing input
- Accuracy with full set
  - FW: 92.7% vs 92.3%
  - DA: 91.4% vs 91.8%
Evaluation

- Motion inference

- Weight updating
  - Available input: Left Arm
  - LA’s weight increases

Accuracy (%)

<table>
<thead>
<tr>
<th>Activity Sets</th>
<th>Free-weight</th>
<th>Daily-activity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ArmTro</td>
<td>MULT</td>
</tr>
<tr>
<td>92.7%</td>
<td>92.3%</td>
<td>91.4%</td>
</tr>
<tr>
<td>91.8%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Weight

<table>
<thead>
<tr>
<th>Iterations</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Arm</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Right Arm</td>
<td>0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Trunk</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Conclusion 1, 2, 3

1. **One goal:**
   - Understanding *human arm motions*

2. **Two aspects:**
   - *Real-time* tracking
   - Motion inference *tolerant to missing inputs*

3. **Three techniques:**
   - HMM state *reorganization*
   - Hierarchical search
   - Attention-based network *adaption*