AWARE@ICU

Colin Lea

Automated Workflow and Activity Recognition
CIS2 Checkpoint Presentation #10
overview
pipeline

Raw Signal → Derived Signals → Action

- Staff Tracking
- Patient Tracking
- Equipment ID
- Gestures
- Binary Events

i.e. “Greeting”
<table>
<thead>
<tr>
<th>Milestone</th>
<th>Mid Feb</th>
<th>End Feb</th>
<th>Early March</th>
<th>Mid March</th>
<th>End March</th>
<th>Early April</th>
<th>Mid April</th>
<th>End April</th>
<th>Early May</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-Gestures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-Locations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-Skeleton</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-Scene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-Windows</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-Recog</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Docs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Recorder

<table>
<thead>
<tr>
<th>Task</th>
<th>Difficulty</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Purpose Recorder</td>
<td>Easy</td>
<td></td>
</tr>
<tr>
<td>Face removal</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Blackout stretcher</td>
<td>Easy</td>
<td></td>
</tr>
<tr>
<td>Anonymization</td>
<td>Medium</td>
<td></td>
</tr>
</tbody>
</table>

## Experimentation

<table>
<thead>
<tr>
<th>Task</th>
<th>Difficulty</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Get IRB Approval</td>
<td>Easy</td>
<td>Waiting</td>
</tr>
<tr>
<td>Record preliminary data</td>
<td>Easy</td>
<td>Non-publishable</td>
</tr>
<tr>
<td>Ensure de-identification</td>
<td>Easy</td>
<td></td>
</tr>
<tr>
<td>Hand annotate activities</td>
<td>Easy</td>
<td>4/4/2012</td>
</tr>
<tr>
<td>Additional recording</td>
<td>Easy</td>
<td></td>
</tr>
</tbody>
</table>

## Derived Data

<table>
<thead>
<tr>
<th>Task</th>
<th>Difficulty</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location tracker</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Face direction tracker</td>
<td>Easy</td>
<td>CSIRO Software</td>
</tr>
<tr>
<td>Staff body pose estimation</td>
<td>Hard</td>
<td>In Process</td>
</tr>
<tr>
<td>Patient tracker</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Equipment Identification</td>
<td>Hard</td>
<td></td>
</tr>
</tbody>
</table>

## Activity recognition

<table>
<thead>
<tr>
<th>Task</th>
<th>Difficulty</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gesture recognition</td>
<td>Hard</td>
<td></td>
</tr>
<tr>
<td>Retrospective analysis</td>
<td>Hard</td>
<td></td>
</tr>
<tr>
<td>Patient Tracking</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Multi-camera models</td>
<td>Medium</td>
<td></td>
</tr>
<tr>
<td>Evaluation</td>
<td>Easy</td>
<td>5</td>
</tr>
</tbody>
</table>
• 3 hours of depth/RGB/skeleton footage
• 2 Kinects
• Pediatric ICU
ICU depth video

Head view

Foot view
kinect skeletons

Camera 1:

(+)=location::skeleton=joints positions::green=correct::red=wrong

Camera 2:
kinect skeletons

Camera 1:

(+)= location :: skeleton = joints positions :: green = correct :: red = wrong

Camera 2:
Can we do better?
activity pipeline

Raw Signal → Skeleton Tracking → Gestures → i.e. “Greeting”

Derived Signals

Patient Tracking

Equipment ID

Binary Events

Action
skeletal tracking

Depth Image → Segmentation → Skeleton
segmentation (step 1)

Depth Image \[ \text{Mean Image} \] \[ \text{Background Subtraction} \]
segmentation (step 2)

**Problem:** people are shades of gray
- Iterative solutions too slow

**Solution:**
- Method 1: Cluster
- Method 2: Gradient/edge removal

(Method 1) Clustering (DBSCAN)  
(Method 2) Gradient-based
segmentation
tracking
skeletal tracking (in progress)

Find extremities (ie. arms, head)
- Floyd-Warshall Algorithm (like Dijkstra’s)
- Think flow along surface!

Parts-based model
- All people have similar structure
- Initialize with head + general direction (SVD)

Flow from centroid

**Gesture Models**

**Action Recognition**

Staff gestures

- Determine sub-actions
  - i.e. Inserting tube
  - Giving medicine

**Algorithm**

1. Calculate PCA per-joint, per-class
   - 1 \([x, y, z]\) basis * 13 joints * n classes
2. Correlate input w/ each known gesture’s PCA basis
3. Find similarity between test and training (per joint, per class)
4. Vote (Optional: Add prior)
dimensionality reduction

My Solution

All gesture poses

Per joint, per class
PCA component

All gesture poses

Per class PCA component

Per class normalization

[Bigdelou ‘12]
experiment

10 gestures
5 trials/gesture
2 people

Wave
Circle (counter-clockwise)
Circle (clockwise)
Push forward
Push left
Push right
Swoosh right
Reach up
Duck
Kick
results

5 samples w/ leave one out

Avg: 94.5% [0.95, 0.95, 0.95, 0.925, 0.95]
Train on one person, test on another

Train A, Test B: 44% [0.425, 0.475, 0.425, 0.4, 0.475]
Train B, Test A: 61% [0.575, 0.5, 0.7, 0.65, 0.625]
One-shot learning

Avg: 77% [0.8, 0.75, 0.75, 0.825, 0.725]
results

Sliding window

At one time step

Avg: 95.5% [0.95, 0.95, 0.975, 0.95, 0.95]
Sequences
other directions

**Action Recognition**
Time-series graphical models
- Switching Linear Dynamical System (code available)
- CRFs

Multiple Instance Learning
- Split into bags -> Iterative SVM
Questions?
All gesture poses → Per class PCA component → Per class normalization
All gesture poses

Per joint, per class
PCA component