Personal Sensing
Continued

Tarek Abdelzaher
Dept. of Computer Science
University of Illinois at Urbana Champaign
Transportation Mode Detection

- Based on phone accelerometers

Challenge: Gravity Estimation

- Accelerometers measure the superposition of two forces:
  - Forces due to acceleration along each axis
  - Forces due to gravity along each axis

- Since accelerometer orientation is unknown, it is hard to separate the two.
  - What is the gravity component of the measurements read along each axis?
Solution #1

- Insight: Gravity is fixed. Other forces come and go.

- Solution: Average the acceleration measurements over a long enough time window
  - Transient forces will tend to cancel out
  - Constants (i.e., gravity) will remain

- Pros/cons?
Solution #2

- If there are intervals of relatively low accelerometer variance, it means that the accelerometer is not “moving”.
  - Acceleration measurements during those intervals are mostly attributed to gravity.
  - Remember current measurements and set them as the “gravity components”
- Reset measured gravity components when significant motion is detected.
Features

- Mean, variance, kurtosis, integral, autocorrelation, zero crossings, energy, entropy, FFT coefficients, etc.
  - Frame-based
  - Peak-based
  - Segment-based
- Standard classifier from prior work
# Features

<table>
<thead>
<tr>
<th>Domain</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical</td>
<td>Mean, STD, Variance, Median, Min, Max, Range, Interquartile range</td>
</tr>
<tr>
<td></td>
<td>Kurtosis, Skewness, RMS</td>
</tr>
<tr>
<td>Time</td>
<td>Integral, Double integral, Auto-Correlation,</td>
</tr>
<tr>
<td></td>
<td>Mean-Crossing Rate</td>
</tr>
<tr>
<td>Frequency</td>
<td>FFT DC,1,2,3,4,5,6 Hz, Spectral Energy, Spectral Entropy, Spectrum peak position,</td>
</tr>
<tr>
<td></td>
<td>Wavelet Entropy, Wavelet Magnitude</td>
</tr>
<tr>
<td>Peak</td>
<td>Volume (AuC), Intensity, Length, Kurtosis, Skewness</td>
</tr>
<tr>
<td>Segment</td>
<td>Variance of peak features (10 features), Peak frequency (2 features),</td>
</tr>
<tr>
<td></td>
<td>Stationary duration, Stationary frequency</td>
</tr>
</tbody>
</table>
Evaluation

- 150 hour of transportation
- 16 individuals
- 4 countries
- Multiple scenarios (walk, train, tram, metro, bus, ...)

![Map of transportation routes](image)
Evaluation Results

- Better accuracy than competition

<table>
<thead>
<tr>
<th></th>
<th>Peaks</th>
<th>Wang</th>
<th>Reddy</th>
<th>Peaks</th>
<th>Wang</th>
<th>Reddy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stationary</td>
<td>96.1 (0.5)</td>
<td>57.3 (4.5)</td>
<td>81.6 (1.0)</td>
<td>70.0 (2.1)</td>
<td>59.5 (2.3)</td>
<td>70.6 (2.9)</td>
</tr>
<tr>
<td>Walk</td>
<td>93.1 (0.1)</td>
<td>87.2 (0.2)</td>
<td>97.7 (0.1)</td>
<td>95.9 (0.1)</td>
<td>89.1 (0.2)</td>
<td>95.9 (0.1)</td>
</tr>
<tr>
<td>Bus</td>
<td>78.2 (4.2)</td>
<td>71.1 (1.4)</td>
<td>67.3 (1.6)</td>
<td>78.0 (3.3)</td>
<td>70.4 (1.4)</td>
<td>86.2 (6.4)</td>
</tr>
<tr>
<td>Train</td>
<td>68.2 (5.0)</td>
<td>32.1 (0.8)</td>
<td>7.7 (4.4)</td>
<td>80.1 (4.0)</td>
<td>31.6 (0.7)</td>
<td>55.4 (11.9)</td>
</tr>
<tr>
<td>Metro</td>
<td>64.5 (5.9)</td>
<td>54.4 (0.6)</td>
<td>70.1 (8.8)</td>
<td>82.0 (2.6)</td>
<td>51.4 (0.9)</td>
<td>56.6 (3.5)</td>
</tr>
<tr>
<td>Tram</td>
<td>84.0 (2.1)</td>
<td>58.1 (0.8)</td>
<td>82.8 (7.5)</td>
<td>86.1 (2.1)</td>
<td>58.2 (0.8)</td>
<td>64.5 (7.0)</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>80.1 (2.9)</strong></td>
<td><strong>60.0 (1.4)</strong></td>
<td><strong>68.0 (3.9)</strong></td>
<td><strong>82.1 (2.4)</strong></td>
<td><strong>60.2 (1.1)</strong></td>
<td><strong>71.6 (5.3)</strong></td>
</tr>
</tbody>
</table>
Evaluation Results

- Confusion matrix

<table>
<thead>
<tr>
<th></th>
<th>Stationary</th>
<th>Walk</th>
<th>Bus</th>
<th>Train</th>
<th>Metro</th>
<th>Tram</th>
</tr>
</thead>
<tbody>
<tr>
<td>62733</td>
<td>755</td>
<td>135</td>
<td>58</td>
<td>302</td>
<td>361</td>
<td></td>
</tr>
<tr>
<td>1549</td>
<td>63664</td>
<td>456</td>
<td>650</td>
<td>723</td>
<td>805</td>
<td></td>
</tr>
<tr>
<td>3976</td>
<td>647</td>
<td>32400</td>
<td>18</td>
<td>104</td>
<td>8118</td>
<td></td>
</tr>
<tr>
<td>6730</td>
<td>330</td>
<td>874</td>
<td>31921</td>
<td>5907</td>
<td>1894</td>
<td></td>
</tr>
<tr>
<td>5057</td>
<td>711</td>
<td>2961</td>
<td>10879</td>
<td>41203</td>
<td>2682</td>
<td></td>
</tr>
<tr>
<td>4341</td>
<td>318</td>
<td>10123</td>
<td>87</td>
<td>2067</td>
<td>77715</td>
<td></td>
</tr>
</tbody>
</table>
Evaluation Results

- Fragmentation: Fraction of ground truth events recognized as multiple events

![Graph showing fragmentation rate of evaluated approaches for different modes of transportation: Stationary, Walk, Bus, Train, Metro, Tram, Car.](image)
Evaluation Results

- **Under-fill**: Percentage of time missing from an event due to detection latency
Indoor/Outdoor Detection

- Possible solution: GPS?

Indoor/Outdoor Detection

- Possible solution: GPS?
Indoor/Outdoor Detection

- Possible solution: IO-Detector
  - Light sensor (higher-intensity outdoors)
  - Cell signal (stronger outdoors)
  - Magnetometer (higher fluctuations indoors)
Supervised Classification

- Sensors: light, proximity, magnetic, microphone, cell, WiFi, GPS, battery thermometer, etc.
- Multiple classifiers selected from WEKA library
- Primary feature set
  - Light intensity, cell strength, magnetic variance
- Extended feature set
  - Light intensity, sound intensity, temperature, magnetic variance, cell strength, proximity
Supervised Classification

- Single-area data set (trained and tested on one area)
- Multiple-area data set (trained on all areas, tested using cross validation)

1. **Primary features**: Light intensity, Cellular signal strength and magnetic variance. (This is analogous to IODetector, but we use cell signal strength instead of its derivative.)
2. **Extended feature set**: light intensity, sound intensity from microphone, temperature from battery thermometer, magnetic variance, cellular signal strength and proximity sensor value.

---

![Accuracy Comparison](chart.png)
Supervised Classification

- Trained on one area and tested on another

Results show inadequate performance
Semi-supervised Classification

- Clustering (with partial labeling)
- Self-learning
- Co-training
Co-training

- Small amount of labeled data is used to train two classifiers. Unlabeled samples are then classified. The output with higher confidence is used to train both classifiers.

<table>
<thead>
<tr>
<th>Naive Bayes based selection</th>
<th>SVM based selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classifier 1</td>
<td>Classifier 1</td>
</tr>
<tr>
<td>light intensity, time of the day, proximity value, battery temperature</td>
<td>cell signal strength, light intensity, time of day, proximity value</td>
</tr>
<tr>
<td>Classifier 2</td>
<td></td>
</tr>
<tr>
<td>sound amplitude, cell signal strength, magnetic variance</td>
<td>battery temperature, sound amplitude, magnetic variance</td>
</tr>
</tbody>
</table>

![Diagram of co-training process](image)
Co-training Performance

- Accuracy gradually improves as new unlabeled data is introduced
Co-training Performance

- Improves accuracy of detection in unknown environments compared to other approaches
Step Detection

- Detects when people step down from sidewalk to street

Step Detection

- Detects phases of walking

![Graph showing pitch variations over time with labels for push-off, swing, heel-strike, and stance phases.]

- Pitch [degree]
- Time [s]
Step Detection

- Detects slope changes across steps

(a) Trough.
(b) Rise.
(c) Fall.
(d) Step Off.
Performance Evaluation

- High detection accuracy