Indoor Localization
I – Introduction and Positioning Algorithms

Petteri Nurmi
About the course

• Advanced course: networking and services subprogramme (also well suited for algorithms, data analytics, and machine learning)
• 4 credit units for passing the course
• Two compulsory project works
  • Can be done in groups or independently
  • No exercises, no exam
• Course page: [http://www.cs.helsinki.fi/courses/582747/2015/s/k/1](http://www.cs.helsinki.fi/courses/582747/2015/s/k/1)
  • All materials and project work instructions available from the webpage
Objectives

- Learn basic principles of indoor localization technologies and algorithms
- Understand the dimensions that can be used to classify and evaluate localization techniques
  - Which signals can be used for measuring locations?
  - Which estimation algorithms can be used?
  - Which kind of errors affect localization?
- Be able to apply knowledge into practice
  - Two project works
Localization

• Refers to determining where an entity is located
  • Also known as *positioning*, *locating*, or *position tracking*
  • Literature separates between three target areas:
    – Indoor, outdoor, and ubiquitous (everywhere)
• Outdoor localization well-established area
  • Satellite positioning (GPS, Glonass, others) used in cars, boats, aircrafts, mobile devices
  • Requires “clear” view of sky hence not available indoors
• Course focuses on localization for *indoor* areas
  • Satellite positioning and other outdoor positioning methods covered as part of the *Location-Awareness* course
Why Indoor Localization?

- Indoor environments becoming increasingly complex (and commercialized)
  - COEX mall (Seoul, Republic of Korea): \(\approx 154,000 \text{m}^2\) with multiple floors
  - Indoor navigation, location-based advertising, etc.
- Indoor logistics and analytics
  - Monitoring people flows, optimizing staff and product placements, identifying bottlenecks in environment etc.
  - Airports, supermarkets, and other indoor locations starting to use localization techniques
What are the main challenges?

- Layouts of environments complex
  - Mixture of open and confined areas
  - Most signals subject to refraction and attenuation
- Varying crowd densities
  - Humans absorb and influence most signals that are used in positioning
- Accuracy-cost tradeoffs
  - Best accuracy requires expensive investments
- Maintenance costs
  - Even small changes in environment can requires efforts at maintaining desired level of accuracy (e.g., recalibration)
Localization Systems

• Localization system responsible for localizing an entity
  • We mainly focus on humans, but could equally be an animal or an object (key, package, other object)
• Three parts to every localization system
  • Localization algorithm: technique that determines the position of the entity from the given measurements
    – Also known as position estimator
  • Location system: collects measurements that can be used to localize the entity
    – Examples: WiFi, ultrasound, GSM, infrared, …
  • Reference system: determines how output of the system relates to geographical area
Components of Reference Systems

1. Coordinate system: how positions are represented
   • Absolute: position uniquely mapped into geographical area (e.g., 60°12′16″N 24°57′46″E for Kumpula campus)
     – Location meaningful within the scope of the coordinate system
   • Relative: position given relative to another object (5km NE of Stockmann)
     – Location meaningful only in a local context

2. Datum: model that defines how coordinates relate to geographic area

3. Projection: mapping that determines how positions can be visualized
Geographic Coordinate Systems

• Systems for representing locations as vector of numbers, called *coordinates*
  • Coordinates refer to position, typically measured either in terms of distances or angles
• Coordinate system consists of three components
  • Origin: intersection of the axes of coordinate system
  • Scale: the subdivision of axes into common units
  • Orientation: direction of axes
Reference Systems for Indoor Positioning

- Reference systems for indoor localization typically defined relative to a floorplan / map
  - One corner (or the center) of a map defined as origin, positions given relative to origin
  - Datum defined as mapping that converts differences in coordinates to physical distances
    - E.g., difference of 0.025 coordinate units equivalent of 1m or similar
  - Projection defined as mapping that relates coordinates within the floorplan / map
    - E.g., pixel distance, how coordinate differences translate to pixels within the floorplan
Reference systems: Indoors vs. Outdoors

- Outdoor reference systems more complex
  - Earth is ellipsoidal so need to take into account curvature of Earth’s surface in distance calculations
  - E.g., WGS-84: World Geodetic System
    - Specifies a reference ellipsoid that approximates the shape of Earth
    - Separate datums for surface (vertical, altitude estimation) and shape (horizontal, position)

- Indoors:
  - Size of area small → no need to consider curvature (usually) of Earth

Source: http://en.wikipedia.org/wiki/Latitude
Tracking

- Refers to monitoring the position of an entity over time
- Errors in individual position estimates can be potentially decreased by considering the evolution of estimates
- Mathematical foundation for tracking given by so-called state-space models
  - Physical model where the “state” of the system evolves over time and is estimated from noisy measurements
  - In the context of location tracking, the state refers to position, and measurements to position estimates
- General form of the model:
  \[
  \begin{align*}
  x_k &= Ax_{k-1} + v \\
  y_k &= Ux_k + w
  \end{align*}
  \]
  State equation
  Measurement equation
Most techniques for tracking build on the principle of Bayesian optimal filtering

- Estimated location a hypothesis
- Probabilistic methods used to update hypothesis given new measurements and prior information

Common techniques include:

- Kalman filter
  - Optimal implementation of Bayes filter when noise and state space Gaussian
- Particle filter
  - Discrete approximation of optimal filter
  - Most common technique nowadays

Details of tracking out-of-scope for the course
Triangulation

- The process of determining location using angle measurements from known reference points
  - Based on geometry of triangles
- Basic idea:
  - Measure angle from two (or more) reference points
  - Each measurement defines a line and the intersection of lines defines the position of the object
  - Distance from reference points can be determined if the distance between the reference points is known
- Angle estimation complex and requires specialized hardware
  - Discussed in next lecture
Triangulation: Example

\[ x + y = l \]
\[ = d / \tan \alpha + d / \tan \beta \]

\[ d = l / (1 / \tan \alpha + 1 / \tan \beta) \]
\[ d = l \sin \alpha \sin \beta / \sin (\alpha + \beta) \]
\[ d = 105 \sin 38 \sin 43 / \sin (38 + 43) \]
\[ = 41.1 \text{ m} \]
Trilateration

- The process of determining locations using distance measurements from known reference points
  - Based on geometry of circles and spheres
  - One of the oldest positioning techniques in the world
    Used, e.g., with GPS and ultrasound
  - The most popular technique for outdoor positioning
- Basic idea:
  - Each measurement defines a circle of uncertainty where the object can be located
  - The position of the object can be determined from the intersection of multiple circles
Trilateration: Example
Measuring Distance – Time-of-Flight

• When signal velocity known, propagation time can be used to estimate distances

• One-way measurements:
  • Beacon sends system time, client compares the time to its own system time when signal is received
  • Requires that clocks are synchronized
  • Basis of satellite navigation (GPS, Galileo, GLONASS)

• Round-trip time:
  • Time for signals to propagate back and forth between client and reference point (e.g., radar)
  • Works well when reference points sufficiently close
Measuring Distance – Radio Propagation Models

- Alternative to time is to use observed radio wave characteristics to estimate distances
  - Attenuation, wave intensity decreases as a function of distance (and other environmental conditions)
  - A radio propagation model is a mathematical formulation that characterizes radio signal variations
- Obstacles and their material have strong influence on signal attenuation
  ➔ Models work best for positioning in obstacle-free environments
- Models depend on frequency and environment type ➔ each model specific to a particular combination
Radio Propagation Models

Log Distance Path Loss Model

- Model that predicts the reduction of signal intensity (i.e., path loss) as it propagates through space

\[ PL = PL_T x_{dBm} - P_{Rx_{dBm}} \]

\[ = PL_0 + 10\gamma \log_{10} \left( \frac{d}{d_0} \right) + X_g \]

- Transmitted power
- Received power
- Length of path
- Path loss at reference distance (in decibels)
- Path loss exponent
- Reference distance
- Fading variable
Radio Propagation Models
Log Distance Path Loss Model

- Path loss exponent $\gamma$ characterizes different environments
  - Free space $\gamma = 2$
  - Outdoors normally $\gamma = 2.5 - 5$, depending on the presence of obstacles
  - Indoors $\gamma = 1.6 - 6$, depending on the presence of obstacles
- Value of $\gamma$ typically determined empirically
  - Choose a reference distance $d_0$
  - Measure path loss at reference distance
  - Estimate $\gamma$ from measurements
Log Distance Path Loss Model – Example

Given a RSS measurement $s = -95$ how far are we from the transmitter?

- $d_0 = 3\text{m}$
- $PL_0 = 50 \text{ dBm}$
- $\gamma = 3.45$
- $Tx = 40 \text{ dBm}$
- Fading ignored

- $PL = 55 \text{ dBm}$
- $\log_{10} d = (PL - PL_0) / (10 \gamma) + \log_{10} d_0$
- $d = 10^{(PL - PL_0) / (10 \gamma) + \log_{10} d_0}$
- $d = 4.188 \text{ meters}$

$$PL_0 + 10\gamma \log_{10} \frac{d}{d_0} + X_g$$
Log Distance Path Loss Model – Example

- Assume path loss at 2m distance equals 12 dBm, what is the (empirical) value of the path loss coefficient?
- \[ \gamma = \frac{PL - PL_0}{10 \times \log_{10} \frac{d}{d_0}} \]
  \[ = \frac{12 - 0.5}{10 \times \log_{10} 2} \]
  \[ \approx 3.32 \]

\( d_0 = 1 \text{m} \)
\( PL_0 = 2 \text{ dBm} \)

Fading ignored

\[ PL_0 + 10\gamma \log_{10} \frac{d}{d_0} + X_g \]
Multilateration

- The use of differences in distance between two (or more) references points to estimate location
  - A variation of lateration
  - Less sensitive to environmental variations than lateration
- Basic idea:
  - Each difference measurement determines a hyperbolic curve along which the object is located
  - Intersection of two (or more) hyperbolic curves determines the location of the object
  - Difference typically measured using difference between arrival times (time-difference of arrival or TDOA)
Multilateration - Example
In practice, distance/angle measurements contain inaccuracies, e.g., due to:

- Attenuation due to obstacles or atmospheric effects
- Signal interference
- Inaccurate synchronization
- Multipath effects

Instead of obtaining accurate circles, lines or hyperbolas, estimates define an error region within which the true distance is assumed to be
Errors – Dilution of Precision

- Intersection of multiple error regions defines a region of uncertainty for the overall position estimate
  - Size of region depends on geometry of reference points
  - Dilution of Precision (DoP) measures the size of the error region
- Overall error thus depends on two factors
  - Error in distance/angle calculations
  - Geometry of reference points
Dead (or deduced) Reckoning

- Estimate position by extrapolation from last position
  - Also referred to as inertial navigation
  - Indoors: Pedestrian Dead Reckoning (PDR)

- Requires information about
  - Direction of motion
    - E.g., compass or gyroscope
  - Velocity of motion or distance travelled since last known position
    - E.g., accelerometers or odometers

- Errors in motion measurements amplify over time ➔ accuracy drifts and decreases over time
  - Needs periodic recalibration

Source:
http://en.wikipedia.org/wiki/Dead_reckoning
Dead Reckoning

- Let \((x_0, y_0)\) denote the current position of the target.
- New position \((x_1, y_1)\) given by
  \[x_1 = x_0 + L \cos \alpha, \quad y_1 = y_0 + L \sin \alpha\]
- Where \(L\) is the distance traveled.
- If distance is not known, but velocity is, we have \(L = v \Delta t\)
Fingerprinting

- Technique that exploits spatial variations in observed signal characteristics for positioning
- Two phases:
  - Calibration: construct a database that contains measurements of signal characteristics at different locations
  - Estimation: compare current measurement against the database and estimate position using the best matches
- Basic idea thus similar to lateration:
  - Distances reflect differences between measurements of signal characteristics instead of physical distance
  - Reference points defined manually in calibration phase instead of using pre-existing points
Fingerprinting

- Fingerprinting requires that signal characteristics are stable and sufficiently distinctive
  - Spatial variation: signals at different locations vary enough to be distinguishable
  - Temporal variation: signals sufficiently stable over time
  - Device variation: differences between measurements of different devices sufficiently small
- Generally *any* signal source can be used as long as its stable enough across these variations
  - WiFi and magnetic field most commonly used (indoors)
  - Other sources include audio and even background radioactivity
Representing Fingerprints

- Beacon-based fingerprints:
  - Identifiers of beacons that can be observed
    - WiFi: access point identifiers (MAC)
    - iBeacon: Bluetooth beacon identifiers

- Signal characteristics-based fingerprints
  - Vectors that capture specific characteristics of the measured signals
    - E.g., WiFi / Bluetooth: signal strength (RSS) of beacons
    - Magnetic positioning: strength of magnetic field components, possibly also orientation/heading of device

- Difference-based fingerprints
  - Vectors that capture differences in signal characteristics
  - Also known as hyperbolic fingerprints
    - WiFi: differences between access point strengths
    - Audio: differences in audio intensity between predefined frequency bands
Examples of fingerprint representations

WiFi scan:

Beacon

00:21:91:52:20:c8, 00:21:91:51:5e:5e, 00:21:91:52:26:bc

Signal characteristics

00:21:91:52:20:c8, 00:21:91:51:5e:5e, 00:21:91:52:26:bc

Hyperbolic

58/55, 58/60, 55/60
Indexing Fingerprints

- Deterministic
  - Positioning based on individual fingerprints
  - Nearest neighbors based position estimation
- Probabilistic (or model-based)
  - Fingerprints used to construct a model which is stored and used for positioning
  - Requires multiple measurements per location
Deterministic Fingerprinting

- Let \( s \) denote a vector consisting of measurements of signal characteristics
  - Pairs consisting of an identifier and RSS value
  - Deterministic fingerprinting calculates the distance \( d(s, x) \) between measurement \( s \) and all fingerprints \( x \) in the database
    - Euclidean distance most common choice
      \[
      d(s, x) = \sqrt{\sum_{i=1}^{N} (s_i - x_i)^2}
      \]
  - Position estimated based on the distances \( d(s, x) \)
Deterministic Fingerprinting

- **kNN (k nearest neighbors)**
  1. Find k best matching measurements in the database
  2. Estimate position as the geometric average of the locations associated with these measurements

- **WkNN (weighted k nearest neighbors)**
  1. Find k best matching measurements in the database
  2. Assign weight for each of these measurements using the difference in signal characteristics
     - Example: inverse of the distance
  3. Estimate position as a weighted centroid
Deterministic Fingerprinting Example (k = 5)

<table>
<thead>
<tr>
<th>Current measurement</th>
<th>Radio map</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:21:91:52:20:c8</td>
<td>2.0, 5.0, 00:21:91:52:20:c8 61</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>2.0, 1.5, 00:21:91:52:20:c8 70</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>3.5, 5.0, 00:21:91:52:20:c8 58</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>2.5, 4.5, 00:21:91:52:20:c8 56</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>4.5, 3.5, 00:21:91:52:20:c8 55</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>4.2, 2.5, 00:21:91:52:20:c8 72</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>3.2, 1.5, 00:21:91:52:20:c8 77</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>1.2, 2.5, 00:21:91:52:20:c8 50</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>2.6, 1.8, 00:21:91:52:20:c8 62</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3.4, 2.5, 00:21:91:52:20:c8 61</td>
<td>3</td>
</tr>
</tbody>
</table>

2.7 3.16
Fingerprint Matching: Probabilistic

- Use a probabilistic model to capture signal variations at different location
  - Histogram-based: model signal variations using a histogram of observations
  - Parametric: use a parametric distribution to model signal variations (e.g., a Gaussian)
- Position can be estimated by
  - Calculating probability or likelihood of different locations
  - Using the location with the highest probability as the estimate or calculating a weighted estimate
Probabilistic Fingerprinting – Histogram cf. Gaussian

Can be converted into probabilities using $n_i / \sum n_i$ for each separate bin/value $i$
Probabilistic: Histogram

- Histogram is a discrete representation of the observed signal values at a particular location.
- Corresponds to conditional probability of observing a particular value at a given location.

To avoid zero probabilities, a small constant can be added to all values.

Alternatively, values can be interpolated.

\[ p(71) \approx 0.118 \]
Assume we observe value $s = 56$, determine whether the client is located at $x$ or $y$.

\[ p_x(56) = 0.08 \]

\[ p_y(56) = 0.0015 \]
Probabilistic Fingerprinting: Gaussian

Assume Gaussian model for signal strength values:

\[
\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{(x - \mu)^2}{2\sigma^2}\right)
\]

<table>
<thead>
<tr>
<th>Room</th>
<th>Mean</th>
<th>St.Dev</th>
<th>N</th>
<th>Measurement</th>
<th>Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>88.33</td>
<td>8.07</td>
<td>12</td>
<td>64</td>
<td>0.000525157553911607</td>
</tr>
<tr>
<td>2</td>
<td>75.25</td>
<td>6.04</td>
<td>10</td>
<td></td>
<td>0.0116081147104146</td>
</tr>
<tr>
<td>3</td>
<td>90.16</td>
<td>9.02</td>
<td>18</td>
<td></td>
<td>0.000659501766423425</td>
</tr>
</tbody>
</table>

Estimated position: room 2
Probabilistic Fingerprinting - Multidimensional Example

- How to calculate probability of a multidimensional measurement?
- How to handle missing entries?
  - Replace missing value with a default value
  - Replace missing probability with a default value

Radio map:

<table>
<thead>
<tr>
<th>MAC</th>
<th>Mean</th>
<th>St.Dev</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:21:91:52:20:c8</td>
<td>-58</td>
<td>12.45</td>
<td>29</td>
</tr>
<tr>
<td>00:21:91:51:5e:5e</td>
<td>-55</td>
<td>11.04</td>
<td>29</td>
</tr>
<tr>
<td>00:21:91:52:26:bc</td>
<td>-60</td>
<td>11.93</td>
<td>29</td>
</tr>
</tbody>
</table>

Fingerprint

<table>
<thead>
<tr>
<th>MAC</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:21:91:52:20:c8</td>
<td>-52</td>
</tr>
<tr>
<td>00:21:91:51:5e:5e</td>
<td>-59</td>
</tr>
</tbody>
</table>
• Probability can be calculated in two ways
  • Using multidimensional probability distributions
    – Take into account correlations between different transmitters / beacons
  • Assuming transmitters/beacons independent, in which case probabilities can be multiplied together
    – Usually calculated on a logarithmic scale $\implies$ becomes a sum of logarithmic probabilities
    – Logarithms also help to prevent underflows

\[
p(s|\mu, \sigma) = \prod_{i=1}^{N} p(s_i|\mu_i, \sigma_i) \iff \log p(s|\mu, \sigma) = \sum_{i=1}^{N} \log p(s_i|\mu_i, \sigma_i)
\]
### Probabilistic Fingerprinting - Multidimensional Example

#### Radio map:

<table>
<thead>
<tr>
<th>MAC</th>
<th>Mean</th>
<th>St.Dev</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:21:91:52:20:c8</td>
<td>-58</td>
<td>12.45</td>
<td>29</td>
</tr>
<tr>
<td>00:21:91:51:5e:5e</td>
<td>-55</td>
<td>11.04</td>
<td>29</td>
</tr>
<tr>
<td>00:21:91:52:26:bc</td>
<td>-60</td>
<td>11.93</td>
<td>29</td>
</tr>
</tbody>
</table>

#### Fingerprint

<table>
<thead>
<tr>
<th>MAC</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:21:91:52:20:c8</td>
<td>-52</td>
</tr>
<tr>
<td>00:21:91:51:5e:5e</td>
<td>-59</td>
</tr>
</tbody>
</table>

\[
\log p(-52 \mid -58, 12.45) = -3.56
\]
\[
\log p(-59 \mid -55, 11.04) = -3.45
\]
\[
\log p(? \mid -60, 11.93) = ?
\]
\[
\log p(-105 \mid -60, 11.93) = -10.512
\]

or

\[
\log 10^{-16} = -36.8414
\]

\[
p(s \mid \mu, \sigma) = -17.5220
\]

or

\[
p(s \mid \mu, \sigma) = -43.8514
\]
Literature

Literature

- Küpper, A., Location-Based Services: Fundamentals and Operation, Wiley, 2005