Mobile Sensing: Spring 2015
Exercise: 3: Solutions

Due on 26th March 2015 by 17:45 PM.

Instructions: All course participants are requested to submit their exercise solutions electronically to the instructors (samuli.hemminki at cs.helsinki.fi and teemu.pulkkinen at cs.helsinki.fi), as well as to the course lecturer (petteri.nurmi at cs.helsinki.fi) by the due date (latest before the exercise session). In all the exercises, do not just give the answer, but also the derivation how you obtained it. Participants are encouraged to write computer programs to derive solutions to some of the given problems.

Ex 1. Load SensorData1 [1]. containing accelerometer measurements.

a) Divide the data into frames of 50 samples each.

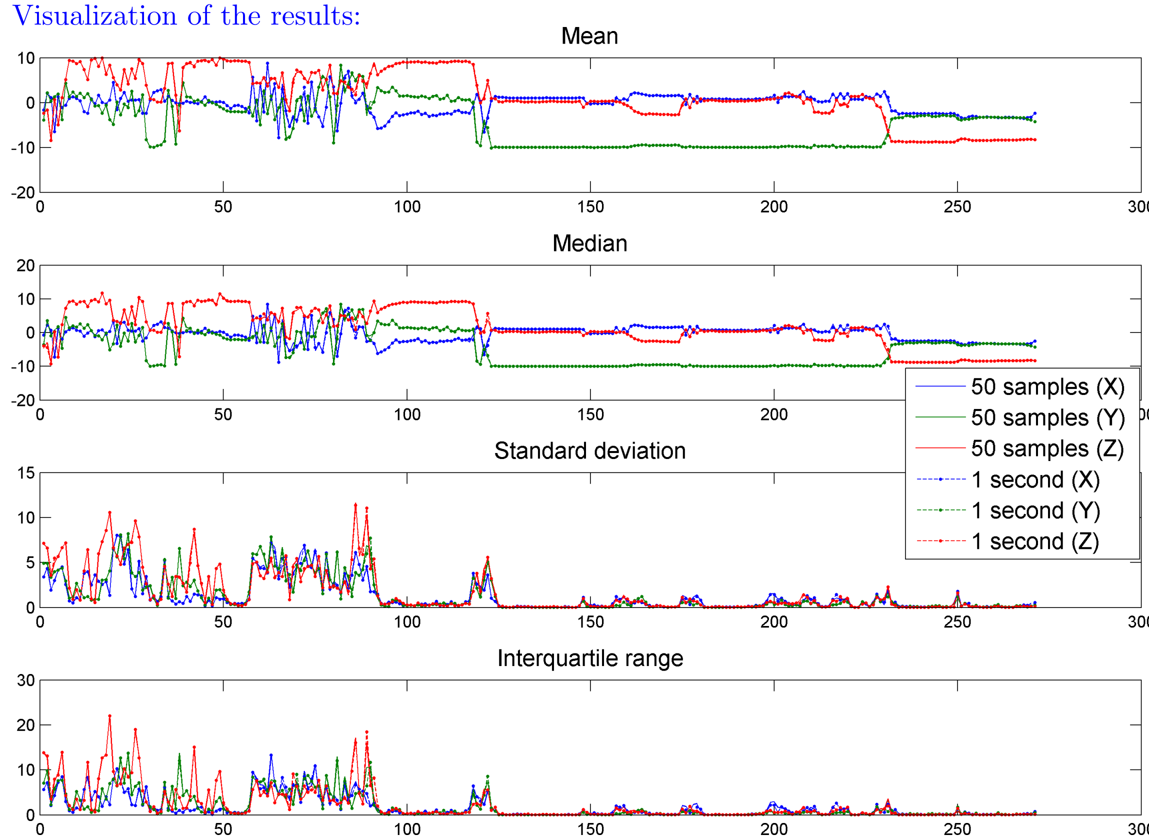
b) Divide the data into frames of 1 second duration each.

c) For both cases a) and b), calculate and compare the mean, median, standard deviation and interquartile range of each frame.

Example MATLAB code for Ex 1-3: Solutions

[1] In the data, columns 1 to 4 corresponds to: [timestamp, x, y, z]

Visualization of the results:



Ex 2. Load SensorData2 [2] containing gyroscope measurements which have been collected simultaneously with the data contained in SensorData1.

a) Perform sample synchronization for the two sensors: SensorData1 and SensorData2.

b) Interpolate sensor values (from both sensors) with uniform 20ms intervals, starting from the timestamp given by the first accelerometer measurement.

Ex 3. Consider a scenario where 4 devices, D_1,D_2,D_3 and D_4, sample their microphones. Table 1 summarizes similarity ρ between the audio signals, Table 2 summarized the information value α of each device and Table 3 summarizes energy costs associated with sensor sampling for each device. Additionally, we define the energy due to data transmission and processing overhead $33mW$ per device. Information gain G of adding device D_i to a set of devices $D_{j...n}$ is calculated using formula: G(D_i,$D_{j...n}$) = (1-max(ρ(D_i,$D_{j...n}$))) * $\alpha(D_i)$

a) Assuming an energy budget of 300mW, write a function which finds the set of devices $D_{i.,.j}$ yielding highest information. What is the achieved information?

[2]In the data, columns 1 to 4 corresponds to: [timestamp, x, y, z]

b) Assuming information requirement of 7 units, find the most energy-efficient device combination.

ρ	D_1	D_2	D_3	D_4
D_1	1	0.5	0.7	0.2
D_2	0.5	1	0.8	0.1
D_3	0.7	0.8	1	0.1
D_4	0.2	0.1	0.1	1

Table 1: Device similarity matrix

α	**Information**
D_1	2.8
D_2	5.5
D_3	6.2
D_4	2.4

Table 2: Information value

mW	**Energy Costs**
D_1	94mW
D_2	87mW
D_3	112mW
D_4	61mW

Table 3: Device energy cost

In addition to the solutions linked above, the solution to this task uses a separate helper function (link).

Ex 4. DTW / Distance

Let's do a little detective work!

In a user study on the movement patterns of customers within a supermarket environment, three location traces were recorded from three different users. In addition, one of these three users carried a cheaper and less accurate device (more noise, less samples). Due to a mixup in the study setup, however, the researchers forgot which user was carrying the cheaper device.

Your task is to figure out which user carried the faulty device. Specifically, implement the Dynamic Time Warping algorithm and compare the different user traces to the mystery

trace. Which user would you suggest carried the device?
Traces are in CSV-format, with columns representing [X,Y]-coordinates. Use the Euclidean distance as the cost function.

User 1

User 2

User 3

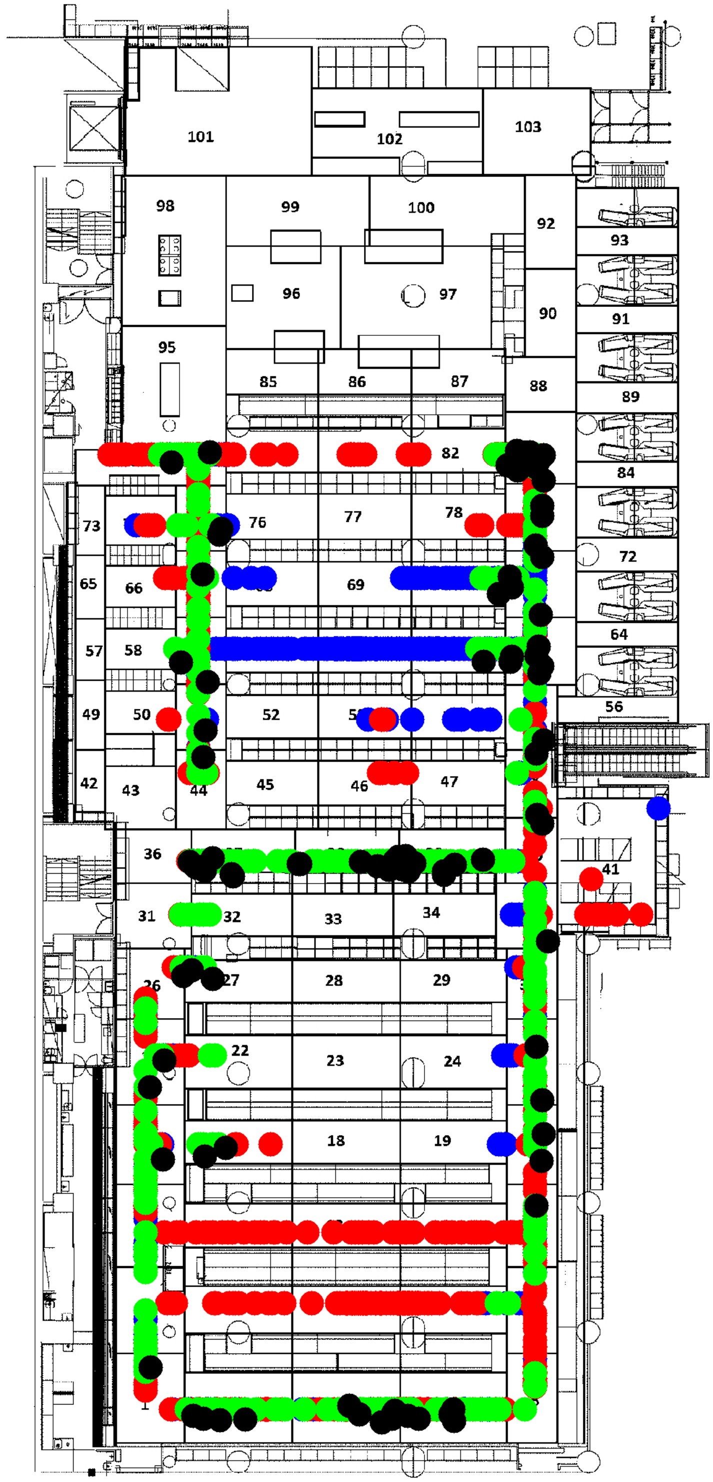
User Mystery

A good "sanity check" for data such as this is to plot the coordinates and inspect the paths visually. In the below picture we have plotted the routes on top of the actual floorplan of the supermarket (User1 = blue, User2 = red, User3 = green, User Mystery = black). From this we can see that the trace of the mystery user most closely resembles that of User 3 (and is in fact the only trace to visit the same aisle, for instance).
Running DTW on the given traces gives us the distances:

- User 1 - User Mystery: $\approx$ 122877
- User 2 - User Mystery: $\approx$ 96016
- User 3 - User Mystery: $\approx$ 13718

Since the DTW distance is the smallest between User 3 and the mystery user (by quite a large margin), this is our best bet on who carried the faulty device.

Example implementation (MATLAB): DTW.m


101
102
103
98
99
100
92
93
96
97
90
91
95
85
86
87
88
89
82
84
73
76
77
78
72
65
66
69
64
57
58
49
50
52
56
42
43
44
45
46
47
36
41
31
32
33
34
26
27
28
29
22
23
24
18
19