Assignment 1a: Collecting and preprocessing sensory data for biking lanes

In this assignment you will collect data and perform basic cleaning, preprocessing and feature extraction. For the data collection you will work according to the data collection protocol below. Follow the steps in the protocol carefully.

Upload the results of your work as a notebook in pdf format before the deadline set on Canvas (currently, September 22nd 20:00).

Protocol

Introduction

Background and relevance

As cities around the world strive to become more bike-friendly, the quality of biking infrastructure has become a critical issue. Smooth and well-maintained biking lanes are essential for the safety and comfort of cyclists. However, many bike lanes suffer from poor maintenance, leading to uneven surfaces with potholes, cracks, bumps, and other imperfections. These issues not only make cycling uncomfortable but also increase the risk of accidents and injuries.

In urban environments, cyclists frequently encounter varying lane conditions that can impact their riding experience and safety. Identifying and addressing these problematic areas is crucial for urban planning and the maintenance of biking infrastructure. However, manual inspection of bike lanes can be time-consuming and resource-intensive. As a result, there is a growing interest in developing automated tools to monitor the condition of bike lanes using technology that is readily available to cyclists.

Smartphones come equipped with sensors such as accelerometers, gyroscopes, and gravity sensors. These sensors can capture data related to the movement and orientation of the cyclist and their bike. By analyzing this sensor data, it is possible to detect whether a bike lane is smooth or bumpy. Automated detection of lane conditions can provide valuable data for city planners and maintenance crews, allowing them to prioritize repairs and improve the overall biking experience.

Aim and Research Question

The aim of this assignment is to gather and preprocess the data required to develop a tool that can automatically determine the condition of bike lanes—whether they are smooth or bumpy. The project will explore the use of sensor technology for data collection and apply basic techniques for cleaning and preprocessing the data. This process involves capturing sensor data from smartphones as cyclists ride over various lane surfaces.

The assignment will focus on several key tasks:

1. **Data Collection:** Utilizing smartphone sensors to gather detailed motion data while cycling on smooth or bumpy biking lanes.
2. **Data Cleaning and Preprocessing:** Applying fundamental techniques to clean and preprocess the collected data, ensuring it is in a suitable format for analysis.
3. **Preparation for Analysis:** merging the datasets, feature extraction and structuring the preprocessed data into a well-organized data frame, ready for the application of clustering algorithms.

Methods

Participants

Groups of 3 are assigned, in each group, 2-3 participants will be recruited .

Inclusion criteria:

- Participant knows how to bike.
- Participant is part of this course.
- In each group, all participants are members of the same assignment group.

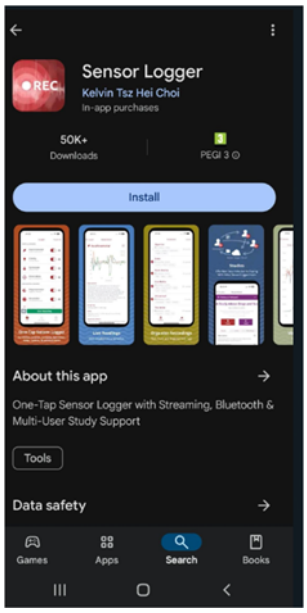
Exclusion criteria:

- Any medical condition (including sports injuries) that does not allow biking.
- Does not know how to bike.

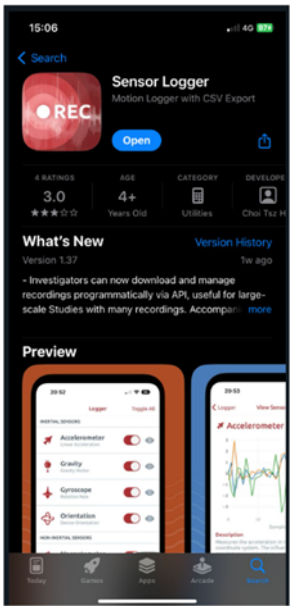
Measurements:

Data will be collected using the free Sensor logger app (available on iOS and Android). Data from the gyroscope, accelerometer and gravity sensor will be used. The other sensors are switched off during data collection. The app shall be used with its default settings and using the free version.

Android



iOS



(For iPhone) If you have issues accessing the (live) calibrated data of the Sensor Logger app, you need to enable motion and compass calibration on your phone as explained here. You will notice this once the live view and the export yield empty results.

Procedures

When selecting a biking lane for data collection, a typical red asphalt cycling lane without imperfections is considered smooth.

When selecting a bumpy biking lane, please choose a lane that exhibits rough or bumpy features but is still commonly used by cyclists. The lane should have some noticeable imperfections, such as small potholes, cracks, or uneven surfaces, but it should not be excessively dangerous or pose significant safety risks. The goal is to capture typical variations in lane quality that cyclists encounter during regular use, not extreme or hazardous conditions.

Make sure the lane you choose is one that you would feel comfortable biking on regularly and that is representative of the kind of surfaces urban cyclists frequently navigate. Your safety is a priority, so avoid any lanes that are in poor repair to the extent that they could be considered unsafe.

Each participant collects data using their own mobile device in the following conditions:

- Bumpy lane: start recording with the sensor, put the phone in your pocket, ride the bike along the lane for about 30 seconds, stop biking, take out your phone and stop recording.
- Smooth lane: start recording with the sensor, put the phone in your pocket, ride the bike along the lane for about 30 seconds, stop biking, take out your phone and stop recording.

Each condition is repeated 5 times by each participant. Make sure that the mobile device is placed in the same pocket in all experiments.

Each participant downloads the data to their personal device and performs their own data cleaning and preprocessing. After cleaning and preprocessing, the datasets of the group members are merged. Then, Feature extraction and data compression is performed and the final data frame is generated.

Data management

Data collection, cleaning and preprocessing is performed on the participant's personal devices. Only processed data is shared without identifiers to other group members. Each participant saves the raw and processed data for (at least) the duration of the course on a personal device. The combined data set is saved by at least one of the group members for the duration of the course on a personal device. The merged dataset is removed after the completion and the grading of the course. The data may only be used for assignment 1b of the course 5ARB0 and may not be shared with other groups or used for other analyses.

Ethical Considerations

In this study only subjects capable of cycling can participate. The movements that have to be made are very low risk and the number of participants in this study is very low. Persons that have higher risk of injuries when performing the cycling movements, because of medical conditions or sports injuries, are excluded from participation to further reduce risks.

This research does not fall under the scope of the WMO. Furthermore, all participants collect and pre-process their own data. Therefore, no formal informed consent procedure is followed. Participants are informed by providing access to the full data collection and analyses protocol beforehand. By collecting and pre-processing data, participants consent participation. By sharing data with the group members, participants consent to data analysis on the processed data.

Privacy

The privacy risk in this study is low. All participants collect data on their personal device. Data will only be shared with other group members after cleaning and preprocessing. The data do

not contain any direct identifiers and no pseudonymization code will be added to the data. The data may not be used for other purposes than this assignment. This way we minimize the amount of data that will be shared between people.

Assignment template

Your work should be submitted in a notebook pdf format along with data, only one group member should submit on Canvas, and the solution should be placed within the dedicated code cells of the notebook.