

# A visualization and analysis approach of cyclist data obtained through sensors

Johnattan D. F. Viana and Gerson V. A. Neto and Igor M. Galdino  
and Antônio M. B. Oliveira and Reinaldo B. Braga and Carina T. Oliveira  
Federal Institute of Education, Science and Technology of Ceará  
IFCE, Ceará, Brazil

**Abstract**—Solutions for smart cities are being created everywhere in the world, using technology to improve urban infrastructure and make urban centers more efficient and better to live. In this way, the proposal of this work is based on the capture of information through sensors in a smart city context. Sensors are coupled on a bicycle and connected to an Arduino and a Mobile Application. After this, the captured data are saved in a cloud database, displayed and analyzed through a Web Application. In this work, our methodology is organized in three main phases: (i) data collection of the surface in which the cyclist is traveling, and the ultrasonic distance sensor, to identify areas of risk based on the proximity of objects from bicycle, (ii) data analysis and data classification, using machine learning concepts and (iii) data visualization, using map views in a Web Application. This methodology allows the identification of injury risk situations to cyclists. The main contributions of this work are surfaces classification with data collected by the accelerometer and ultrasonic sensor generating useful information through simple data. Real experiments were conducted at Fortaleza (Ceará, Brazil) and Aracati (Ceará, Brazil). This work brings new perspectives to collaborative data collection for identification of injury risk situations to cyclists, since it can be used to suggest routes based on these risk indicators and offer a secure environment for cyclists.

**Keywords**—Cyclists, Smart Cities, Mobile Application, Web Application, Machine Learning.

## I. INTRODUCTION

Nowadays technological advances are part of daily human routines. These advances have become so pervasive that they are often indispensable for simple everyday activities [1]. Besides the creation of tools to bring comfort, safety, convenience and sustainability [2], these technological advances applied in the context of Smart Cities result in urban infrastructure improvement and make urban centers more efficient and better to live [3].

Recently, some of these tools are being designed and developed for the construction of roads and types of transport [4], [5], including bicycles, that are occupying the streets of cities. According to the WorldWatch Institute, in 1965, world production of cars and bikes was essentially the same, with approximately 20 millions units each. However, in the few last years, bike production had climbed to over 100 million units per year compared with 42 millions units of cars. Last year, bicycle production was 127 millions units globally, a 18% increase over 2004. This consumption trend is socially desirable as it has a positive impact on population health and urban traffic [6].

The growth in the number of bicycles consequently led to the increase of exclusive routes for them. However, it does not guarantee the complete safety of its users, either by the lack of attention of drivers or because of holes caused by natural degradation or human actions. Thus, it is useful to have collaborative ways to inform users about possible changes in their daily routes, or to allow the use of new routes to the same destination. In this sense, bicycles must evolve and adapt to this new reality.

The objective of this work is to present a visualization and analysis approach of cyclist data obtained through sensors coupled to a bicycle. In addition, two examples of tools are presented: the first uses data from an accelerometer connected to an Arduino to identify the surface where the cyclist is performing the ride and the second uses data obtained by an ultrasonic distance sensor to signal in a Web Application geographical locations with high risk of accident.

This work is structured as follows. Section II shows researches related to the proposal of this work. Section III details the proposal, as well as the components of its architecture and applications, explaining the process of data capturing and data geovisualization. Results and discussions are shown in Section IV. In Section V, conclusions of this work are described, as well as directions for future works.

## II. RELATED WORK

Studies in Europe and North America show that bicycles are regaining their importance as a means of transportation [7]. Thus, it has been known that the information obtained in traffic context allows the appearance of many studies. For instance, there is an approach that proposes a methodology to estimate and map bicycle volumes and cyclist injury risk throughout the entire network of road segments and intersections on the island of Montreal, achieved by combining smartphone GPS traces and count data [8].

In order to aid in the daily activities of cyclists, there are a variety of existing technologies, for example: COBI<sup>1</sup>, SmartHalo<sup>2</sup> and LIVALL<sup>3</sup>. COBI basically connects smartphones to bicycles, providing assistance to phone charging, remote control, voice guidance, weather forecasts, lighting and anti-theft system. SmartHalo proposes to find the quickest and safest routes for cyclists. Users can inform the destination on their smartphone, keep it in their pocket and follow the light

---

<sup>1</sup><https://cobi.bike/>

<sup>2</sup><https://www.smarthalo.bike/>

<sup>3</sup><https://www.livall.com/>

indications on the bicycle handlebar. LIVALL is helping in cyclists safety when they are on the road, using functionalities as Heart Rate Monitor, Turning signal, SOS Alert and Walkie-Talkie.

In addition, there is also a framework entitled SensorWeb-Bike (SWB) [9]. SWB is a web-based information service designed to support urban environmental monitoring. It is composed by i) a Arduino-based mobile platforms, (ii) an Urban GeoDatabase, and (iii) a Web Application, that enables users to view and analyze the data of volunteered geographic information gathered by cyclists. As proposed in this present work, SWB proposal uses an Arduino-based mobile platform with sensors installed on bikes. However, the type of data collected by the sensors in this platform is focused on monitoring urban air quality and weather parameters, different from the data collected in this work.

As regards to surface classification, the experiments conducted by Abulizi et al. [10] were carried out to detect irregularities in surfaces and to calculate their quality according to international standard International Roughness Index (IRI) but without the use of software for calculation of standards. In other approaches related to Zigbee GPS modules and Z axis accelerometers, Matlab<sup>®</sup> tool was used for calculations according to the model proposed in [11].

The route calculation functionality for the BeCity application combines information provided by the user and information of the routes of the city to calculate the best path, taking into account the distance of the trip, path availability and security [12]. These works share a common objective, which is to inform the routes of the bicycles. However, the main difference is that the algorithm of BeCity focuses on taking the routes by user, and our work will focus on picking the routes based on indicators of risk, traffic and track quality.

### III. PROPOSAL

The proposal of this work is based on the capture of information through sensors that are coupled to a bicycle. These data are saved in a cloud database, displayed and analyzed through a Web Application. Figure 1 details the proposal structuring, which is based on the DeMaiS [13], characterizing the architecture used in this work as compatible with the architectures of other applications in the scenario of Internet of Things and Smart Cities [14], [3].

#### A. Obtaining Data

In this work, two sensors were used: the accelerometer, to collect data from the surface in which the cyclist is riding a bike, and the ultrasonic distance sensor, to identify areas of risk based on the proximity of objects from the bicycle, which can represent traffic information. The data of these two sensors coupled to the Arduino (Accelerometer and ultrasonic) and the Global Positioning System - GPS (the Smartphone itself) are stored in an Online database through a direct connection of Android with MySQL.

1) *GPS Sensor*: A GPS is a satellite navigation system that works from a sensor on a mobile device. The GPS is a technology capable to locate and determine latitude and longitude by using some of the 24 satellites that surround

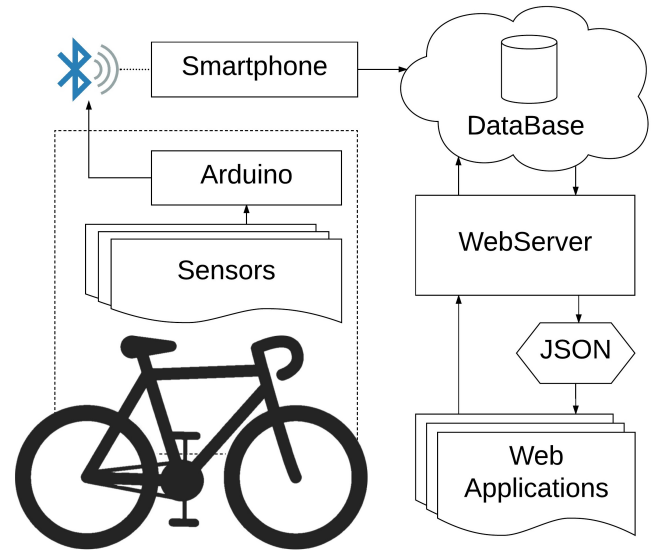


Fig. 1: Diagram Proposal.

Earth. Everywhere in the world, GPS captures information from a group of four of these satellites and can determine its exact location on the map.

2) *Accelerometer Sensor*: MPU-6050 sensor contains an accelerometer and a gyroscope. There are 3 axes for the accelerometer and 3 axes for the gyroscope, being in all 6 Degrees Of Freedom (6DOF). In addition to these two sensors, there is a built-in feature called DMP (Digital Motion Processor), responsible for making complex calculations with sensors and whose data can be used, for example, for gesture recognition, navigation (GPS) systems, games and several other applications. This sensor is highly accurate due to the 16-bit digital analog converter for each channel. Therefore, the sensor captures the X, Y and Z channels at the same time. We chose the use of an external Accelerometer because each cellular model has its own board, resulting in different vibratory levels because of hardware heterogeneity.

3) *Ultrasonic Sensor*: Our approach uses an ultrasonic distance sensor on the back of the bike. This sensor used in the solution is the HC-SR04, which is capable of measuring distances of 2cm ( $d_{min}$ ) to 4m ( $d_{max}$ ) with a precision of 3mm. Even being used only for short distances, this sensor was sufficient for the tests performed in this work. In our experiments we considered an average cyclist speed of 20 km/h, without stop point, that is caused for example by traffic lights. The operation of the HC-SR04 is based on the emission of ultrasonic signals sent by it, which waits for the return of the signal (*echo*) and, based on the time between sending and returning, calculates the distance between the sensor and the detected object. This procedure is represented by Equation 1.

$$d = \frac{echo * V_{sound}}{2} \quad (1)$$

In Equation 1,  $d$  represents the distance that the identified object is from the sensor, which is obtained by multiplying the value of *echo* and speed of sound, divided by 2. The speed of sound can be considered ideally equal to 340 m/s, then the



Fig. 2: Surface classification: collected data from sensors

result is obtained in meters considering the time in seconds. In the equation, the division by 2 is due to the fact that the wave is sent and received back. Therefore, it travels 2 times the desired distance.

### B. Data Geovisualization

To the data visualization, the proposal uses a webserver that performs a request to the DB and returns a processed and transformed response into JavaScript Object Notation (JSON), as previously shown in the Figure 1. The JSON format was used because it is possible to structure the information in a more compact and readable way, thus making it easier to analyze this information.

In relation to the development of maps, it was used the open code JavaScript library called Leaflet. Leaflet allows to build interactive maps in Web projects in a simple way, with high performance and usability, joining in just one library all the necessary tools for the creation of geospatial applications. The cartographic images used in the maps for Leaflet were from Open-StreetMap (OSM), a Free/Libre Open Source Software (FLOSS).

After the transformation of the data into JSON, followed by the reception of the data in the JSON format by the page, the information is represented on the map. This representation has the objective of generating information enriched from data coming from the sensors, allowing georeferenced contents, which provide potential analysis. These data require scalable methods of analysis because they need to consider the particular characteristics of the geographic space, such as

heterogeneity, diversity of characteristics and relationships, as well as spatio-temporal similarities.

## IV. RESULTS AND DISCUSSIONS

In this section, we present the results of the evaluation. We focus on the most interesting results that show the behavior our solution in a real scenario of smart city. By means of the methodology utilized in this work, it is possible to attribute and refer each data obtained by Accelerometer Sensor and Ultrasonic Sensor to the latitude and longitude corresponding to the acquired geographical position by GPS Sensor. Table I shows an example of collected data structure used in this work. It stands out that although the data are collected together at the same time, the visualization and analysis of these data are performed separately, for different purposes. The first one uses data from the fourth column of the Table I, referring the vibratory parameters to classify surfaces (Section IV-A). The second one uses the last column of the Table I to map risky locations based on distance parameters (Section IV-B).

### A. Surface classification

In this topic was evaluated the possibility of classifying different types of surfaces through vibratory data collected by an accelerometer fitted on the bike. Thus, a field data survey was done initially, to later perform the categorization of the surface in asphalt and pavement, applying Machine Learning techniques.

For the data collection, was used an Arduino UNO board, a MPU-6050 Accelerometer and a Bluetooth HC-05 module, all powered by a 9V battery. This set was mounted inside



TABLE I: Example of Collected Data Structure

Date	Hour	Geographic Position	Vibratory Parameters	Distance Parameters
dd-MM-yyyy	HH:mm:ss	GPS Sensor Data	Accelerometer Sensor Data	Ultrasonic Sensor Data

a special bag on the bicycle top tube. For the classification, studies were carried out on the technologies that could help in the categorization of the data collected. Machine Learning concepts and Neural Networks were the ones closest to the expected categorization of the data, since each one of them provides a range of classifiers that help in the proposed work.

Matlab<sup>®</sup> has several classifiers that allow various tests to define which one is most appropriate for the data obtained. Table II shows a comparison between the results of tests applied to data used in this work, indicating the Minimum, Maximum and Average percentages of each classifier.

Three classifiers were tested: *SimpleTrees*, *KNN* and *BaggedTrees*. *SimpleTrees* are basically classification and regression trees that predict responses to data, following the decisions in the tree from the root (beginning) node down to a leaf node in reason of predict a response. The *KNN* uses a set  $X$  of  $n$  points and a distance function to find the  $k$  closest points in  $X$  to a query point or set of points  $Y$ . The *BaggedTrees* fit many large trees to bootstrap-resampled versions of the training data, and classify by majority vote.

As shown in Table II, the first and second have great performance, different from the third one that uses too much memory, taking 10 times more run the classification tests. To achieve faster results, the classifier *SimpleTrees* was chosen for this work because it showed the best results in a shorter implementation time frame. Decision trees do not require many tweaks, and with the help of the right tool they can be easily configured to bring fast results.

For the initial analysis using *SimpleTrees* classifier, three filters were applied: Average, Standard Deviation (*SDv*) and Principal Component Analysis (*PCA*). Table III shows the Hit Percentage applying several different combinations of these

filters using this classifier. It was noticed that Average Values filter has better Hit Percentage when applied, either individually or in conjunction with the other filters. For this reason, *SimpleTree* classifier was used with *SDv* filter.

The field surveys were carried out in a route totaling 4.7km of asphalt and pavement surfaces, as shown in Figure 2. Were collected approximately 54,000 points, generated for data modeling and testing. The data collection was divided into:

- 10 rides, 1 minute gathering each, surface asphalt;
- 10 rides, 1 minute gathering each, surface paved Stones;
- 1 ride of nine minutes, with timed time to determine when the ride was over surfaces asphalt and paved stones, to be used to simulate the classification after data training.

The data training followed the steps: Data modeling, Training and Validation. The data modeling consists of assembling a structure that transforms the data into valid information for classification. In analysis, it was verified that with no delay applied, the accelerometer collects approximately 3,100 points per minute, or, 51.6 points per second. The window shown in Figure 3 displays an example from data taken by the accelerometer sensor in 2.5 seconds.

TABLE II: Comparison between Classifiers

Classifier	Min	Max	Average
<i>SimpleTrees</i>	88.61%	98.73%	95.07%
<i>KNN</i>	87.34%	97.47%	93.46%
<i>BaggedTrees</i>	91.52%	99.16%	95.06%

TABLE III: Filter Combinations using *SimpleTrees*

Filter	Hit Percentage
<i>SDv</i> + Average + <i>PCA</i>	87.8%
<i>SDv</i> + Average	87.7%
<i>SDv</i> + <i>PCA</i>	87.0%
Average + <i>PCA</i>	72.0%
<i>SDv</i>	87.5%
Average	70.8%
<i>PCA</i>	68.6%

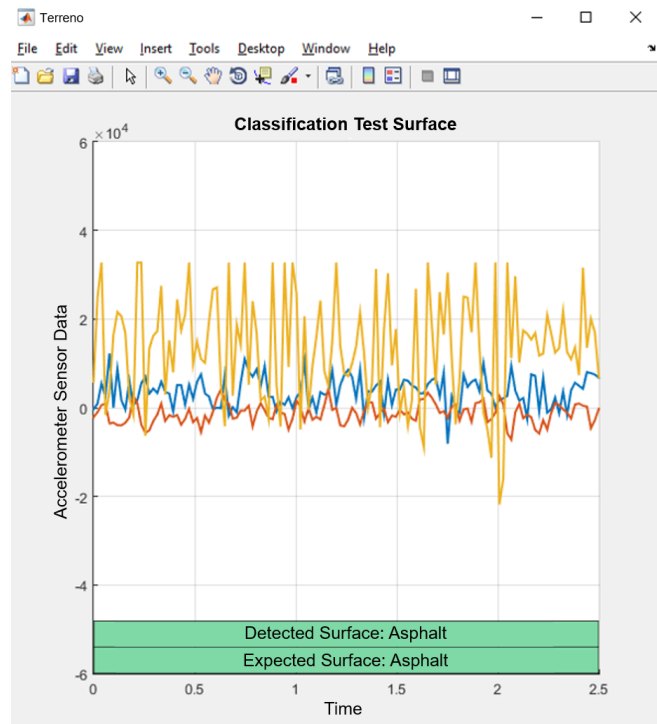


Fig. 3: Example of classifier analysis window

It was used as a base the study done in [15] that determines 2.56 seconds to consolidate a movement. On the rides it was found that this is a valid time interval to consider the moment between the entry of the bicycle into one or more holes and its exit back to the plain asphalt surface. The training is applied to the modeled data, which is then tested. We performed sets of tests that totaled 500 repetitions, which are performed separately in batteries of 50 executions that generate 10 averages of percentage accuracy.

After the latest improvements, an average hit of 94.7% was achieved. To improve this percentage, a method of rejection should be applied in the next version of the code because, according to Bishop [16], it allows to determine a rejection criterion that will minimize the misclassification rate, or more generally the expected loss, for a given fraction of rejected data points.

### B. Mapping risky locations based on the Ultrasonic distance sensor

For these experiments was used an Android mobile application that received, via Bluetooth, data from an ultrasonic sensor connected to an Arduino. Devices used to test the application are LG L70 and LG L90, which have respectively Android operating system in KitKat (4.4.2) and Lollipop (5.0.2). Sensor data collection is performed during the user's ride. Users can view in real time values read by sensor through mobile application. This visualization indicates to cyclists that they should be alert if there is any vehicle approaching the rear of the bicycle. In this work we define values received by sensor to be considered as indicators of risk distances: (i) high (Figure 4a), (ii) medium high (Figure 4b), (iii) medium low and (iv) low.

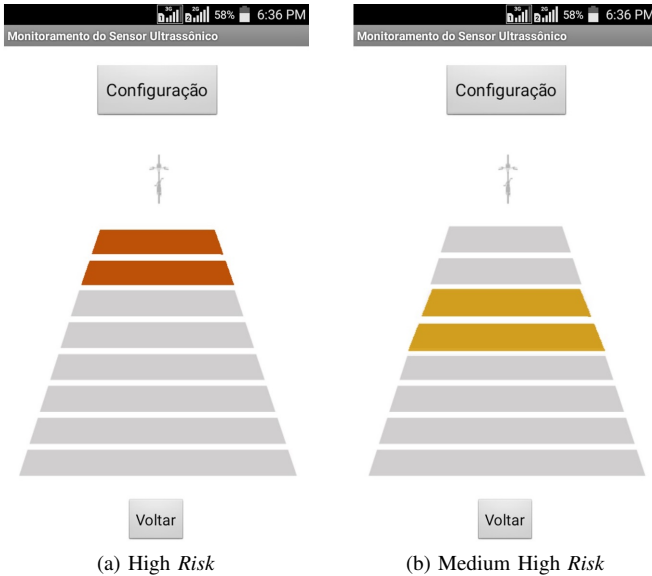


Fig. 4: Risk representation in Mobile Application

In equation 2 these values are represented respectively by the range of  $d_{min}$  to  $a$ , the range of  $a$  to  $b$ , the range of  $b$  to  $c$ , and the interval of  $c$  to  $d_{max}$ , where  $d_{min}$  represents the minimum value read by the sensor and  $d_{max}$  represents the maximum distance captured by it. To perform the experiments,



Fig. 5: Examples of plots in the map view

we have used, respectively, the values of 1m, 2m and 3m for the values of  $a$ ,  $b$  and  $c$ . So, we have:

$$f(d) = \begin{cases} \text{if } d_{min} \leq d < a, & \text{high Risk} \\ \text{if } a \leq d < b, & \text{medium high Risk} \\ \text{if } b \leq d < c, & \text{medium low Risk} \\ \text{if } c \leq d \leq d_{max}, & \text{low Risk} \end{cases} \quad (2)$$

The function  $f(d)$  receives as a parameter the value of  $d$  acquired through Equation 1. With an empirical analysis we defined that the more the distance indicators show that an object is closer to the sensor, the higher the risk situation that the cyclist is exposed. These indicators are significant for interpreting traffic and injury risk information.

The data of sensors are saved in a Database with their respective geographical positions. The risk analysis is performed in the Web application with the object of later signaling risk routes and suggesting better trajectories. Thereafter, data are plotted in the map view, as shown in the Figure 5. Thus, on the map are plotted points of risk referring to proximity information that can refer to a large traffic, and because of that represent an injury risk for cyclist.

Figure 5 shows an example of map plots on the geographical positions where the ultrasonic sensor identified a high risk. Furthermore, if this map has a large number of plots, data can be displayed using a heat map. These approaches of map view can be used, for example, to a route suggestion.

## V. CONCLUSIONS

This paper presents proposal for visualization and analysis of data obtained through sensors, which makes use of a Web application, a Mobile application and sensors in an Arduino

coupled to a bicycle. The results presented demonstrates that it is possible to categorize the surfaces as proposed, even with simpler classifiers, besides highlighting the use of proximity information to signal areas of risk on a map. As a future work, we intend to study other classifiers and define the best to categorize the data, which involve in addition to riding on asphalt and pavement surfaces, the detection of holes in the road. In addition, we plan to ally the tools presented in this work to a set of algorithms [17], and thus suggest routes based on risk indicators, user profiles and user preferences.

#### ACKNOWLEDGMENT

Authors would like to thank IFCE and CNPq that financed this work through funding announcement PAPEX/PROEXT/2016 and PIBITI/PRPI/2016.

#### REFERENCES

- [1] A. Mulay, *Sustaining Moore's Law: Uncertainty Leading to a Certainty of IoT Revolution*. Morgan and Claypool, 2015. [Online]. Available: <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7302711>
- [2] M. R. Alam, M. B. I. Reaz, and M. A. M. Ali, "A Review of Smart Homes - Past, Present, and Future," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1190–1203, 2012.
- [3] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, Feb 2014.
- [4] L. Zhuhadar, E. Thrasher, S. Marklin, and P. O. Pablos, "The next wave of innovation — review of smart cities intelligent operation systems," *Computers in Human Behavior*, vol. 66, pp. 273 – 281, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0747563216306574>
- [5] S. E. Bibri and J. Krogstie, "On the social shaping dimensions of smart sustainable cities: A study in science, technology, and society," *Sustainable Cities and Society*, pp. –, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2210670716305881>
- [6] F. S. D. and L. K., "Cycling for transport and public health: a systematic review of the effect of the environment on cycling," *The European Journal of Public Health*, vol. 21, no. 6, 2011.
- [7] S. Gössling, "Urban transport transitions: Copenhagen, city of cyclists," *Journal of Transport Geography*, vol. 33, pp. 196 – 206, 2013. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0966692313002111>
- [8] J. Strauss, L. F. Miranda-Moreno, and P. Morency, "Mapping cyclist activity and injury risk in a network combining smartphone {GPS} data and bicycle counts," *Accident Analysis and Prevention*, vol. 83, pp. 132 – 142, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0001457515300233>
- [9] C. Vagnoli, F. Martelli, T. D. Filippis, S. D. Lonardo, B. Gioli, G. Gualtieri, A. Matese, L. Rocchi, P. Toscano, and A. Zaldei, "The sensorwebbike for air quality monitoring in a smart city," in *IET Conference on Future Intelligent Cities*, Dec 2014, pp. 1–4.
- [10] N. Abulizi, A. Kawamura, K. Tomiyama, and S. Fujita, "Measuring and evaluating of road roughness conditions with a compact road profiler and arcgis," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 3, no. 5, pp. 398 – 411, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2095756416301994>
- [11] K. Tomiyama, A. Kawamura, S. Nakajima, T. Ishida, and M. Jomoto, "A mobile profilometer for road surface monitoring by use of accelerometers," *7th Symposium on Pavement Surface Characteristics: SURF*, 2012.
- [12] S. Torres, F. Lalanne, G. del Canto, F. Morales, J. Bustos-Jimenez, and P. Reyes, "Becity: sensing and sensibility on urban cycling for smarter cities," in *2015 34th International Conference of the Chilean Computer Science Society (SCCC)*, Nov 2015, pp. 1–4.
- [13] J. D. F. Viana, R. B. Braga, and C. T. Oliveira, "DeMAis: a Decision-Making solution to enhance Accessibility of deaf cyclists in urban traffic Scenarios." *4th International Workshop on ADVANCES in ICT Infrastructures and Services*, 2015.
- [14] M. C. Domingo, "An overview of the internet of things for people with disabilities," *Journal of Network and Computer Applications*, vol. 35, no. 2, pp. 584–596, 2012, simulation and Testbeds. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1084804511002025>
- [15] S. Prasanna, "Machine Learning Made Easy," <http://www.mathworks.com/matlabcentral/fileexchange/50232-machine-learning-made-easy>, 2016, accessed: 2017-05-20.
- [16] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [17] R. B. Braga, "LIDU : Location-based approach to IDentify similar interests between Users in social networks," Doctorate, Université de Grenoble, Oct. 2012. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-00771457>