Scalable Data Set Distillation

Philipp Rigoll, Patrick Petersen,
Jacob Langner, Lennart Ries, Eric Sax | FZI

Data Storage, Analysis & Discoverability

Context data

With large amounts of data it is necessary to facilitate efficient searchability and filtering. One approach is to rely on context data. This enables easy automation and is efficient due to the simple nature of context data. In addition, it is possible to evaluate the coverage of the data without taking the raw sensor data into account.

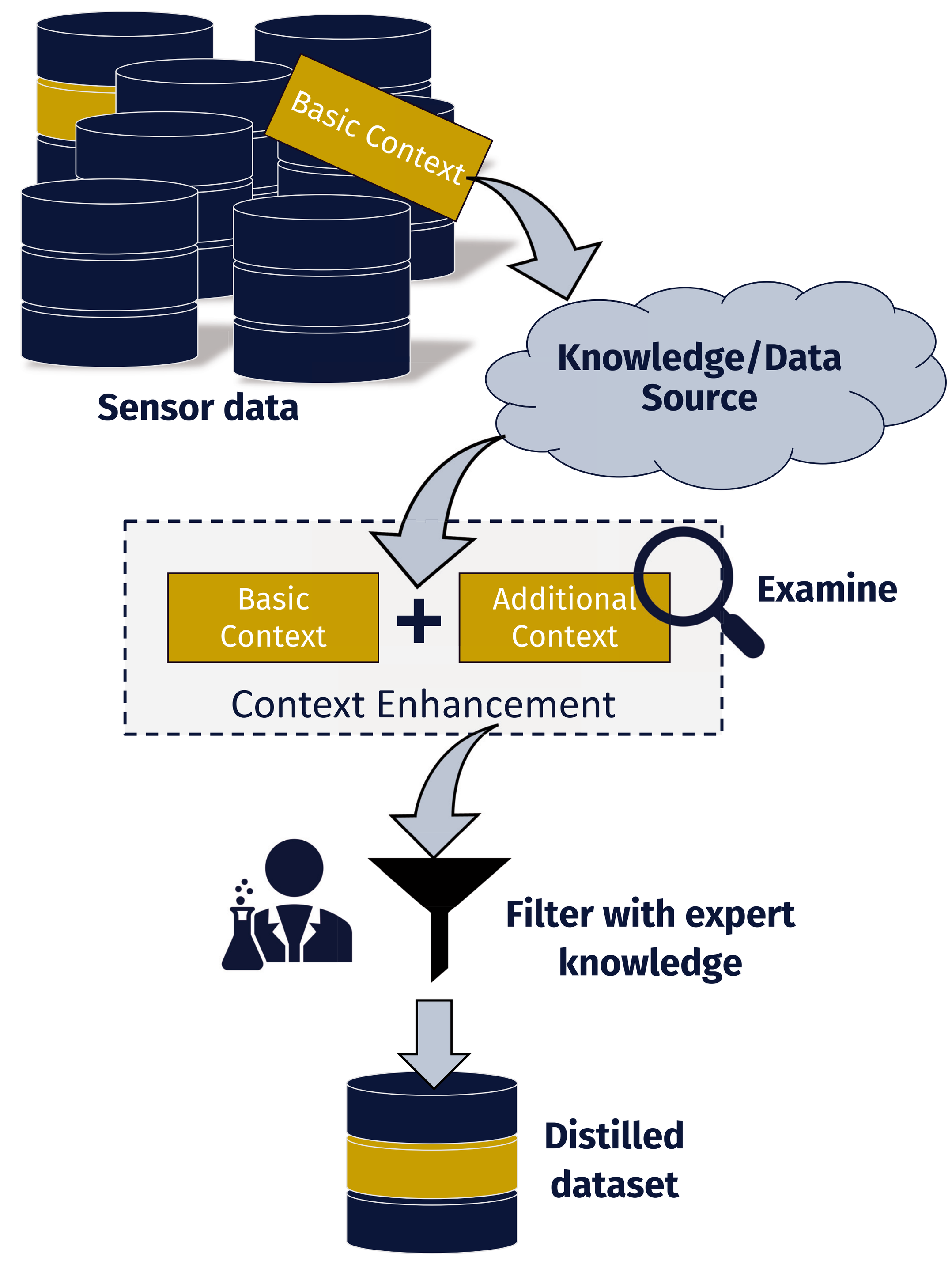

Figure 1: Scheme of the distillation process based on context data [1]

Basic context data

By basic contexts, we mean the data that is recorded as metadata in data acquisition anyway. In our case this basic context data is: time, geographical position and heading. To get an overview of the recorded data the first analyses can be performed on these basic context data points.

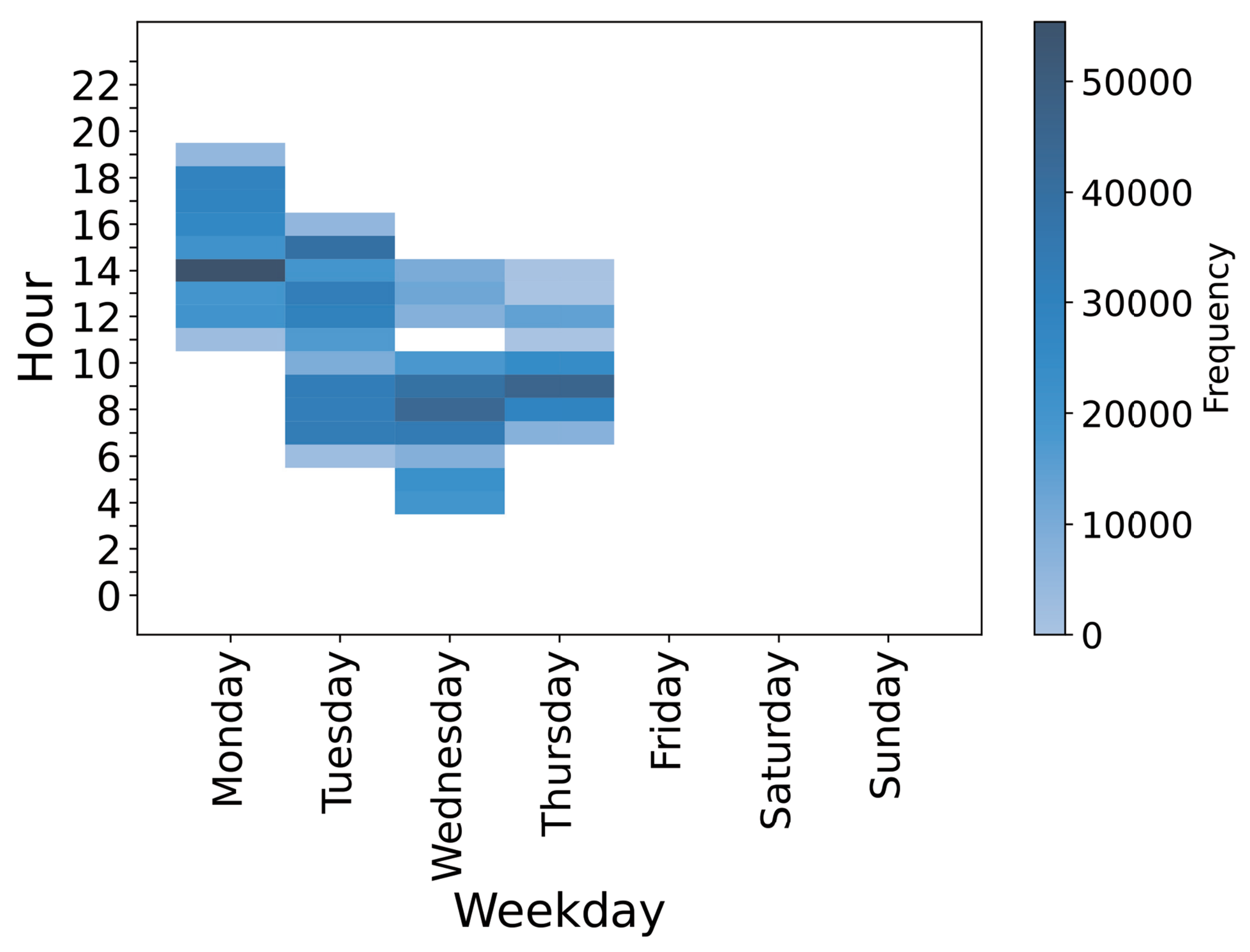

Figure 2: Recording times for a part of the project data

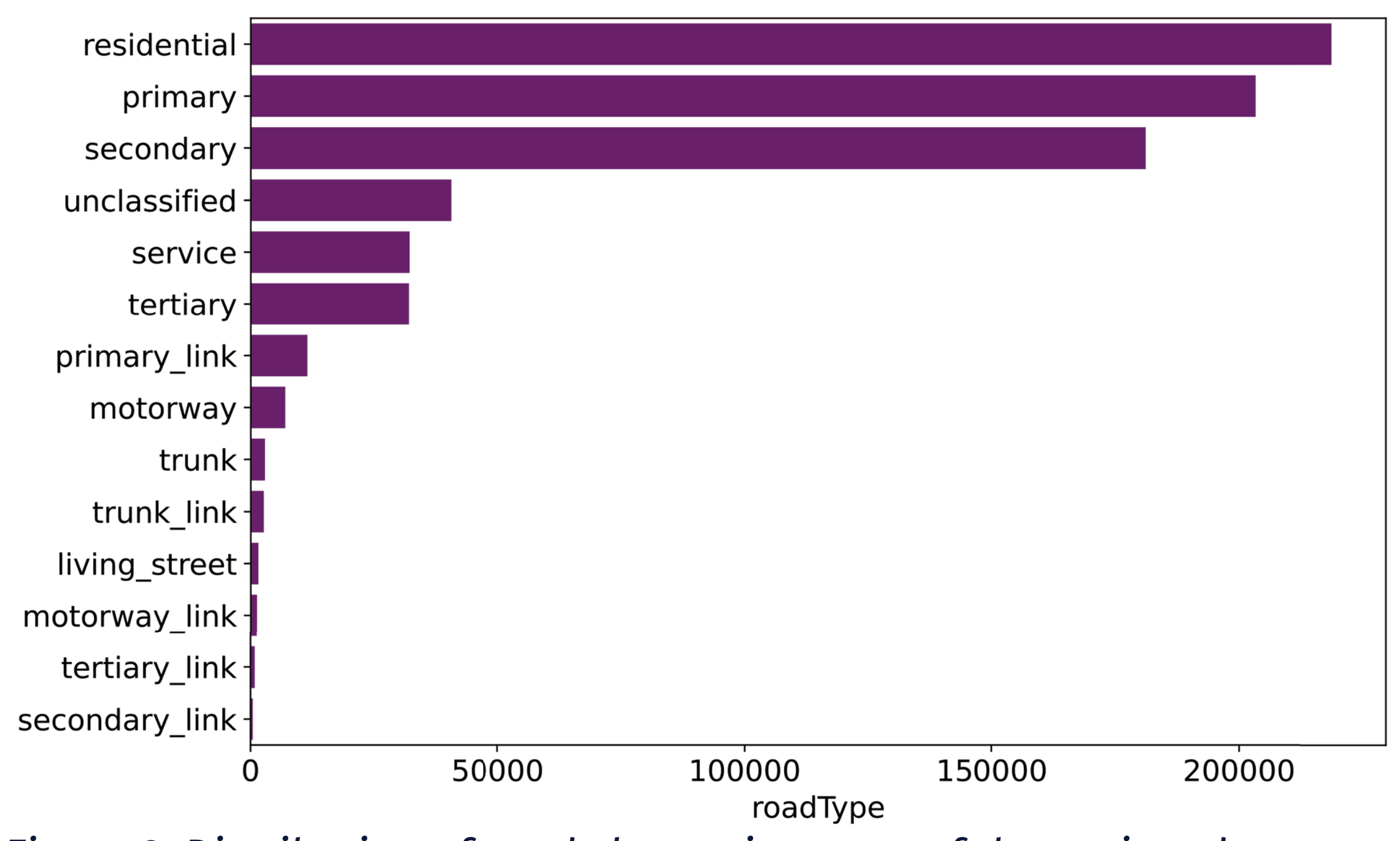

Figure 3: Distribution of road classes in a part of the project data

Enhancement with additional context

Even though these basic contexts provide an impression of the data coverage, they are not sufficient for all use cases e.g. for the composition of special perception data sets. Therefore, enrichment with additional data sources such as geographical data is necessary.

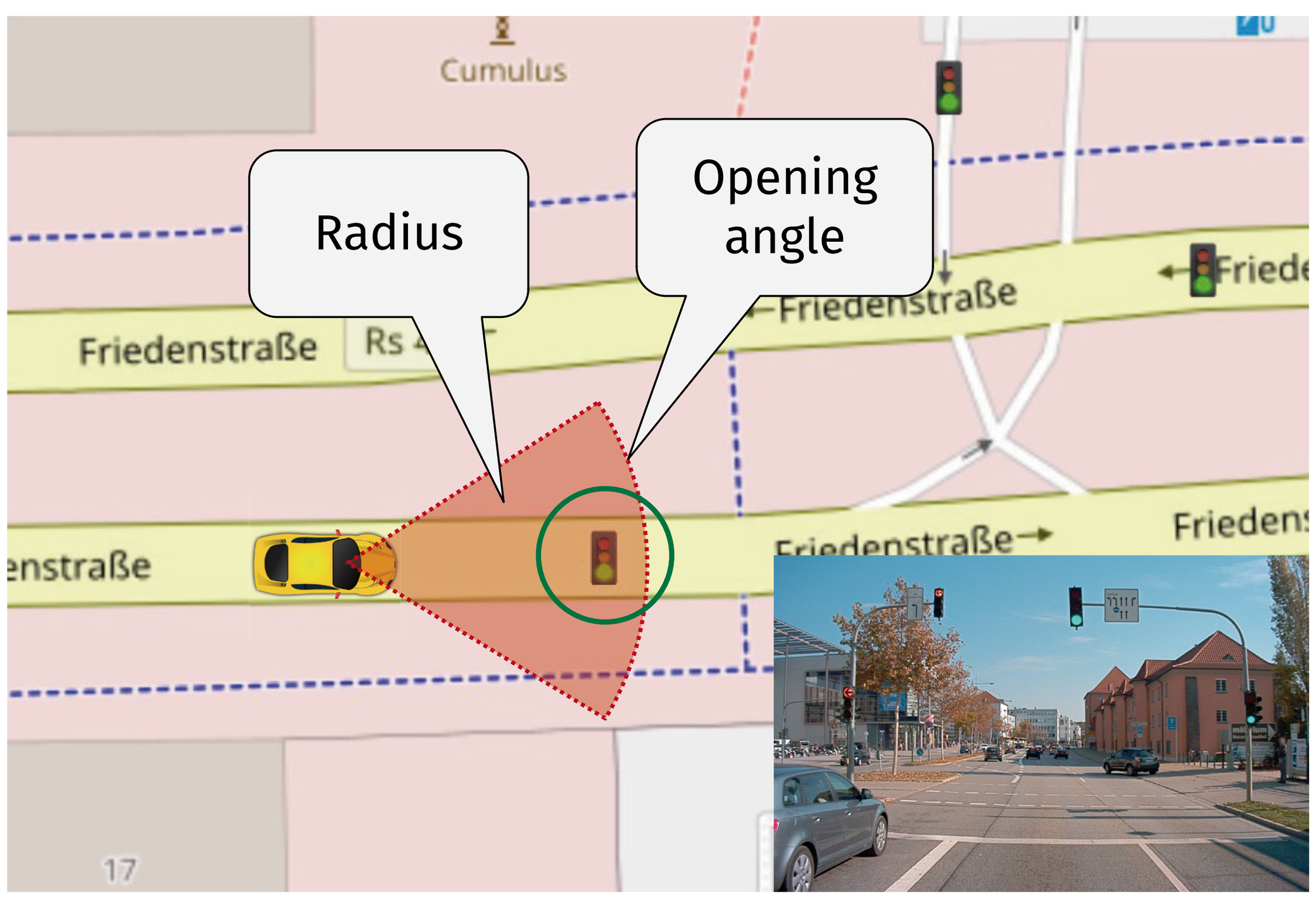

Figure 4: Example that shows for an image from the A2D2 dataset [2] how OpenStreetMap [3] data can be used to check if a traffic light is visible in the image

Based on this advanced context, it is possible to identify special data points such as challenging exposure situations as they occur under bridges. These situations are hard to describe and identify by looking at the sensor data. In contrast, the analysis and filtering on the context level is comprehensible and fast.

Figure 5: Examples of challenging exposure situations in the A2D2 data set [2] which where find with the presented method

References:

[1] Rigoll, Philipp, Ries, Lennart, and Sax, Eric. "Scalable Data Set Distillation for the Development of Automated Driving Functions." 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC). IEEE, 2022.
[2] Geyer, Jakob, et al. "A2d2: Audi autonomous driving dataset." arXiv preprint arXiv:2004.06320 (2020).
[3] Map data copyrighted OpenStreetMap contributors and available from https://www.openstreetmap.org

Partners

External partners

For more information contact:
philipp.rigoll@fzi.de

KI Data Tooling is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

Supported by: Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag