TrimBot2020 Deliverable D8.2

Data Management Plan and Framework

Principal Author: UEDIN
Contributors: UEDIN, RUG
Dissemination: PU

Abstract: The data plan covers the types, formats and quantities of the data, the metadata, and relevant standards. It ensures reliable data analysis and RIA interpretation. This will increase the overall impact of the project results including any future re-use of the data. The data will be stored for at least 10 years after the project on a web server at UEDIN. We do not plan to commercialize the data.

Deliverable due: Month 6

1 TrimBot Garden Dataset

The core data that is planned to be collected and disseminated by TrimBot2020 will be organized in three categories listed below. We are not aware of any similar dataset publicly available. The data will be collected in the purpose-built outdoor test garden located at a project partner facility.

Besides scientific papers the dataset will be accurately described on the website of the project (http://www.trimbot2020.org) and software for data analysis and performance assessment will be provided, together with appropriate documentation for their usage by third-party users.

1.1 Vehicle sequences

We will capture stereo video sequences from the multiple camera systems, synchronised, as the vehicle approaches different scene contexts. Associated laser 3D data 'ground-truth' will also be acquired and semantically labeled. Only selected frames in the video sequence ($\sim$10 fps) will be semantically annotated (every 10th).

1.2 Arm sequences

Similarly we will capture stereo video sequences from the arm camera system as it servos towards trimming positions and trims a bush or plant.

1.3 Garden maps

Additionally we will present 3D SLAM-acquired garden maps, semantically labeled.

2 Standards and metadata

The data format of the dataset follows standards established by community managing Robotic Operating System (ROS), which is also the implementation platform chosen for the project.

The video capturing module will produce a sequence of 'messages' (ROS msg format, http://wiki.ros.org/msg) containing the image data. Message metadata consist of a time stamp and unique identifier (default ROS Header).

Additional modules will process this data further and produce intermediate results, such as depth maps and point clouds, which might be included in the dataset. The concrete extent has to be decided. Data fields in all messages and data types will be specified in detail in Deliverable 3.1 (*Data representation design and implementation*).

The dataset will be recorded and published in the form of ROS 'bagfile' (as described at http://wiki.ros.org/Bags), which is essentially a pack of messages. It can be downloaded by other researchers and played back to simulate the same environment as at the time of capture.

Ground-truth (GT) data will be initially provided in the same format as intermediate results/messages on the same topic, i.e. semantic labels for image segments and surface meshes. This will be specified in detail in Deliverable 7.4 and enable convenient evaluation within the

same framework. We will also export labels and meshes to neutral data formats more widely recognized. Image labels will be stored as indexed bitmaps (PNG file format) and a surface mesh as a list of vertices and faces with label property (PLY file format).

3 Data sharing

We will make use of the institutional DataShare hosting at UEDIN, School of Informatics collection, for public releases of data associated with project publications and other outputs. The service has been recently enhanced to handle large amounts of data expected to be generated by the project (10 GB files, potentially hundreds of GB in total).

We will also deploy an internal GitLab server/group to enable collaboration within the consortium, accessible at `gitlab.inf.ed.ac.uk/TrimBot2020` (subject to change).

Selected components (both data and code) will be mirrored to public GitHub.com server `http://github.com/TrimBot2020` (subject to change). This will allow to link the project results from existing ROS website (`http://github.com/ros`).

4 Archiving and preservation

For final publication of our datasets we will use the DataShare service hosted by UEDIN, which will be listed in the IPAB Collection at `http://datashare.is.ed.ac.uk/handle/10283/291`.

This facility provides unlimited storage and lifetime for research outputs, including backup and versioning. Related policies are described in [1] .

5 Licensing

Our general policy is the data can be freely used by academic users for research. Use of data for commercial purposes will be decided by the partner producing specific data. We consider to use either permissive BSD license or open-source GPL. Specific license will be chosen latest at the time of publication of the first dataset and this document will be updated accordingly.

6 Updates

This initial version of the data management plan will be updated during lifetime of the project to match actual development and publications.

References

[1] University of Edinburgh Information Services. *DataShare Preservation Policy*, 2015. `http://www.ed.ac.uk/information-services/research-support/data-library/data-repository/service-policies/preservation-policy` 4

