Report Proof of concept in Collaboratec

Multimodal 3D imaging of work targets

Client	Work machine industry in Pirkanmaa region
Client main contact	Antti Sirén, antti.siren@fima.fi
CIVIT/Collaboratec main contact	Jussi Rantala, jussi.rantala@tuni.fi, +358503185833

Problem statement

Commercial machine vision solutions for determining work targets for grabbing and handling are mainly based on 2D cameras, optimized lighting conditions and image matching, such as having a top-down view of a conveyer belt with objects at a known distance. In contrast, heavy work machines operate in environments, which are not and cannot be designed for optimal machine vision – the terrain and lighting are constantly changing, and the camera must be placed on the machine itself, limiting the available viewpoints. This creates a need to capture as much information as possible, prompting the use of 3D measurement and capturing the scene in wavelengths outside of the visible spectrum of light.

The objective of this proof of concept (PoC) was to perform dataset collection for the purpose of detection and 3D measurement of targets (work pieces, cargo etc.) for target specific operations in the following use cases:

- Logs littered on the ground. Task: select place to grab and lift.
- Rectangular boxes of cargo on a pallet. Task: determine individual boxes for lifting.
- Shipping container. Task: determine the number of containers and their positions.

This PoC was carried out as a part of the Collaboratec project funded by the European Regional Development Fund (ERDF). CIVIT has conducted a similar project for FIMA in 2019. The differences between the earlier project and the current PoC will be detailed below.

This report describes the used hardware, software and dataset collection process. In addition, sample data is shared and information on how to utilize the collected dataset is given. This dataset collection process can be offered as a service for companies who are interested in customized datasets for their needs.

Scope and objectives

The PoC project investigated the potential of different sensing techniques for 3D imaging of various work targets. This dataset targets the heavy work machines domain and addresses their need to detect and measure objects for safety and autonomous applications. Heavy machines such as crane, forestry machines, port gantry cranes, etc. operate in environments that are not designed for optimal machine vision. The terrain (occlusion by environment or moving objects) and lighting (brightly lit, dark, shadows, light direction) are constantly changing and the cameras must be placed on the machine itself (see Figure 1 for examples).

(a) Night working conditions (b) Occlusion from leaves (c) Shadow on the object

Figure 1: Examples of challenging conditions in work machine industries.

This PoC explored sensors relevant for detecting and measuring work targets in these challenging conditions. The dataset was collected by capturing test scenes along a preplanned route, postprocessing the data and storing in an accessible format. The captures were done in test grounds intended for mobile work machines on the Hervanta campus of Tampere University.

The main results from the PoC are:

- Open sample dataset containing data from the test scenes captured with different sensors
  - Synchronized and calibrated temporal data from all the sensors
  - Post-processed data in easily readable formats for C++, Python and MATLAB
- Report of the work
  - Specifications of the 3D imaging sensors (accuracy, resolution, working range)
  - Definition of the capturing process, which can later be offered as a service for company-specific use cases

Report Proof of concept in Collaboratec

Motivation

Traditional RGB based imaging solutions are not able to address the challenging conditions presented in Figure 1. Instead, it is better to utilize other technology such as depth sensors and hyperspectral sensors. Depth sensors can estimate the depth of objects in front of it. For example, LIDAR sensors shoot rays and based on the time it takes for the rays to return, it estimates the distance of objects in front of it as shown in Figure 2.

Figure 2: LIDAR sensor mounted on a tractor and a sample point cloud example from this setup (Images taken from https://ouster.com/).

Hyperspectral sensors can capture wavelengths outside the visible spectrum. As illustrated in Figure 3, short wave infrared (SWIR) cameras are able to distinguish plastic and metal pieces from beans which are not observable in the visible light.

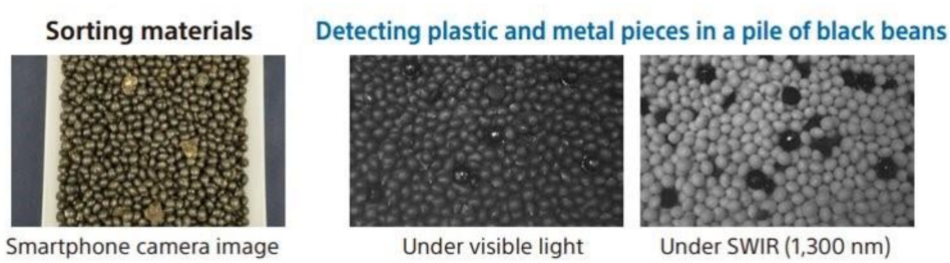


Figure 3: Image of black beans taken with smartphone camera (on the left), with an industrial camera that senses visible light (in the middle) and short-wave infrared light (on the right). (Images taken from https://www.sony-semicon.com/en/products/is/industry/swir.html)

CIVIT conducted an earlier PoC in 2019, in which the suitability and feasibility of using 3D and hyperspectral sensors for 3D detection and measurement was investigated. This PoC contained only static captures of objects. In the first example below (Figure 4), a comparison of LIDAR (Velodyne VLP-16) and stereo camera for detection of a container at 5 to 15m distance was done against the ground truth data from laser scanner (FARO Laser scanner). Though the LIDAR data was sparse it was good enough to fit the container model to this point cloud. On the other hand, the disparity map from the stereo pair had artifacts in the middle which hindered the fitting of the container model.

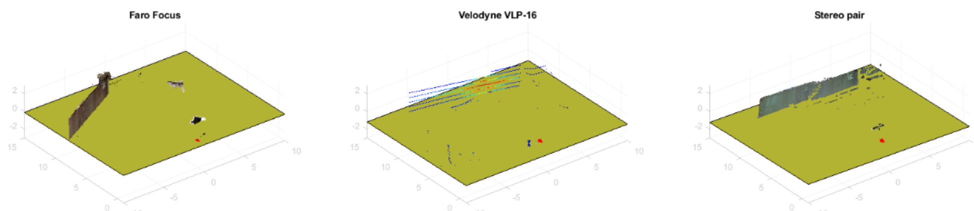


Figure 4: The comparison of ground truth data from FARO Laser scanner in comparison to point cloud from LIDAR data and disparity map from stereo cameras.

The second example is of using hyperspectral data to segment pellets as shown in Figures 5 and 6. One specific spectral band is selected and used for segmentation algorithm. The result of algorithm-based segmentation is comparable to manual segmentation.

Figure 5: Different spectral band of pellets

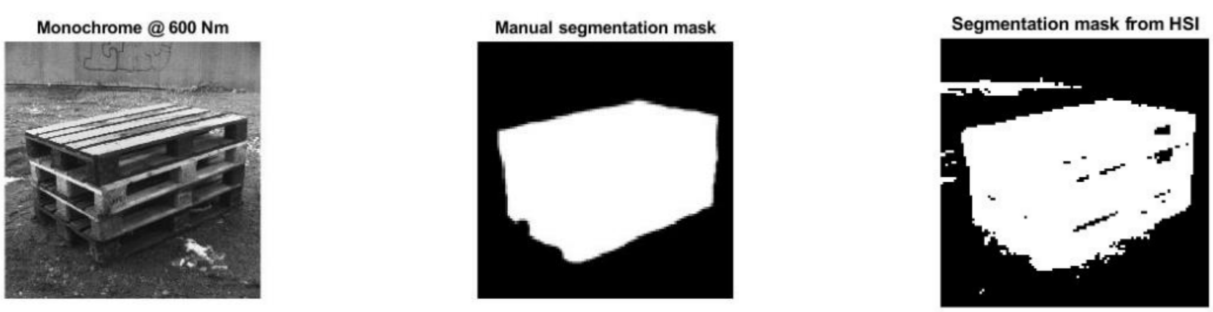


Figure 6: A certain spectral band is selected and used for segmentation. This algorithmic segmentation is compared with manual segmentation.

The current work reported in this document differed from the work done in 2019 in several ways. First, we used different sensors such as new stereo cameras, Time of Flight cameras, short wave infrared cameras, and GNSS+IMU for positioning. Second, the equipment was placed on an Avant wheel loader. Third, continuous data was collected while the vehicle was moving and handling target objects. The aim was to simulate practical work scenarios.

The details of the hardware and software system, dataset collection process and samples from the dataset are presented in the following sections.

System description

The first step was identifying the different types of sensors to be utilized for this dataset collection. Software for recording from these sensors was created for the purpose of dataset collection. The sensors were synchronized to ensure temporally accurate data. The details of the hardware and software implementation are described in detail below.

Sensor information

The sensors shown in Figure 7 were chosen for this PoC. The aim was to include sensors of different type to collect a rich dataset with as much visual information as possible. Details of the sensors are given in Table 1.

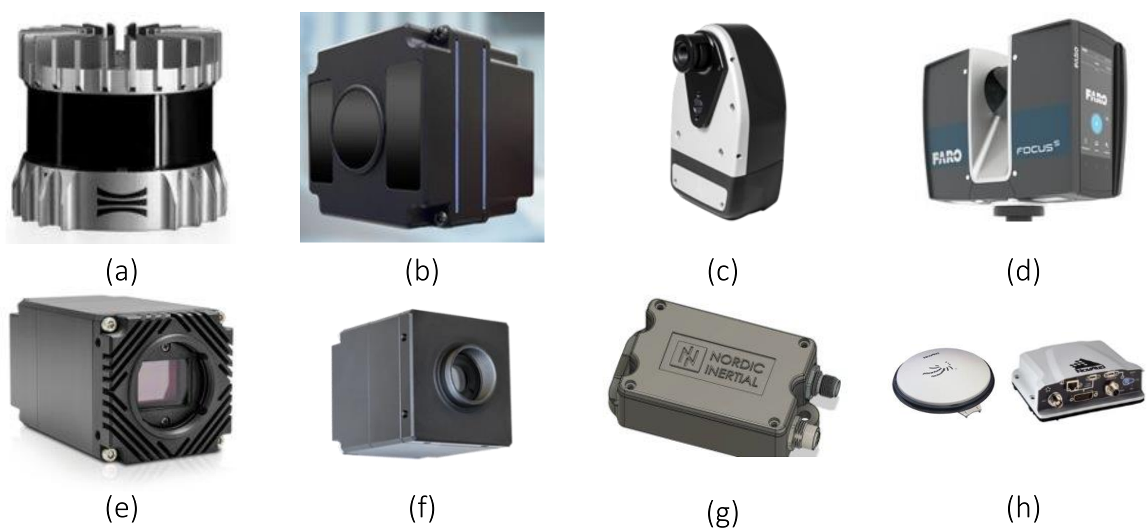


Figure 7: Sensors profiled in the dataset collection included Ouster LIDAR OS0 (a), OToBrite oToCAM500 ToF (b), Senop Rikola HSC-2 (c), FARO Laser scanner (d), Lucid Vision IMX990 SWIR (e), Lucid Vision IMX490 (f), Nordic Inertial IMU (g) and NovAtel GNSS (h).

Table 1: Specifications of the sensors used.

Sensor name	Resolution	Accuracy	Measuring range	Sampling rate
Ouster LIDAR OS0	128 channels 90 vertical field of view 5.2M points per second	±0.9 cm for ranges between 1 – 10 m. Beyond 10 m, precision decays geometrically out to a maximum of ±5 cm at 25 m.	Measures up to 100 m	20 Hz
OToBrite oToCAM500 ToF	640 x 480 pixels 62° x 47°	0.5% of depth, mm level precision	0.3m~7.2m	15 fps
Senop Rikola HSC-2	1024 x 1024 pixels 36.8° x 36.8°		6-18 nm 1000 spectral bands	149 fps
FARO Laser Scanner	360° X 300°	+/-1mm	130 m	5-45
Lucid Vision IMX990 SWIR	1280 x 1024 pixels		400 nm to 1700 nm	84.9 fps
Lucid Vision IMX490 RGB	5.4 MP, 2880 x 1860 pixels			20.8 fps
NovAtel GNSS		~1 cm accuracy		1 Hz
IMU				200 Hz

The FARO Laser scanner and NovAtel GNSS with two antennas were meant primarily for collecting the ground truth of the test environment. FARO Laser scanner was used to capture the ground truth data of the environment. The test area was scanned from multiple locations and the collected data was combined into a single point cloud file.

After post process, NovAtel GNSS provided cm level of accuracy for localization in 1 Hz. In future datasets, in between frames could be accurately interpolated up to 200 Hz with help of the IMU unit. All the sensors except for the GNSS and FARO Laser scanner were connected to a computer that ran the data capture software. The computer had Intel i7-10700TE CPU @ 2.00GHz CPU, NVIDIA GeForce GTX 1660 GPU and 32GB RAM.

Sensor rig

All the sensors were attached to a custom sensor rig (refer to Figure 8). This was done to keep the sensors' relative positions consistent, easily calculate relative positions between them, and make it faster to attach the sensors on a vehicle. The LIDAR was positioned in the middle of the rig. The two RGB cameras were kept on either side of the LIDAR sensor to create stereo depth data. Primary GNSS antenna was attached close behind the LIDAR. Second NovAtel GNSS antenna was not mounted directly on the rig but on the roof rack on top of the vehicle. The computer with other utility equipment like lights were attached to the roof rack as well.

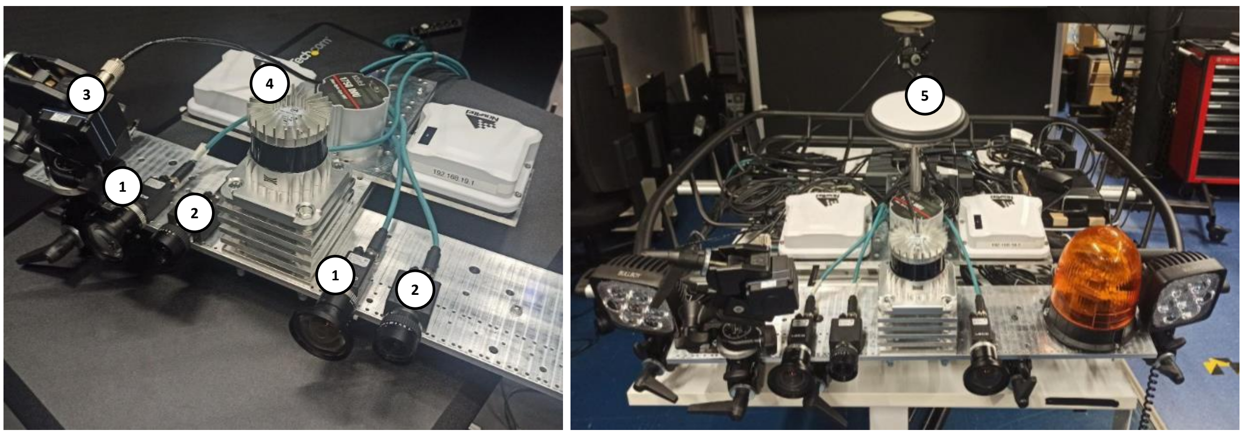


Figure 8: Sensor rig constructed for the dataset collection. 1 = Lucid Vision IMX490 SWIR, 2 = Lucid Vision IMX990 SWIR, 3 = OToBrite oToCAM500 ToF, 4 = Ouster LIDAR OS0, 5 = NovAtel GNSS.

The sensor rig was first mounted on roof rack and then on the Avant wheel loader (see Figure 9). It was kept in place using straps and harnesses. The sensor rig was created in such a way that it could be mounted on any vehicle such as a larger Volvo wheel loader.

Figure 9: The sensor rig mounted on Avant wheel loader.

Data interfacing

The sensors were connected to the computer using ethernet cables, except for the IMU which was connected through CAN cable. The FARO Laser scanner had to be operated through its dedicated user interface, so it was not connected to the computer. The data interfacing is illustrated in Figure 10.

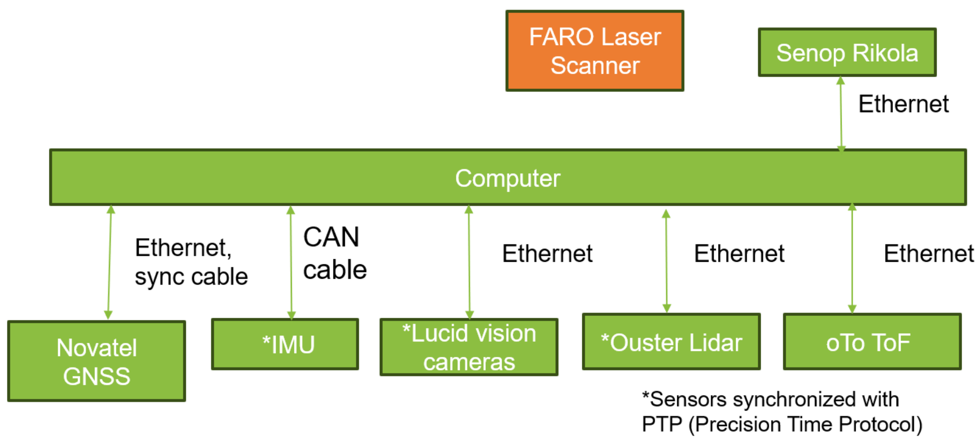


Figure 10: Data interfacing between the different sensors and the computer.

PTP based synchronization

Most of the sensors supported precise time protocol (PTP) for clock synchronization. PTP synchronization was implemented with the ptp4l package. GNSS was the master clock in the system. The computer worked as a boundary clock, receiving PTP time from the GNSS and distributing it forward to the Ouster LIDAR and Lucid Vision sensors. This way, each image from these sensors had synchronized timestamps. IMU and ToF camera did not support PTP so data from them were saved with a timestamp from the PC instead. As the PC was in PTP time, these frames were synchronized as well. This introduced a small delay as the timestamps did not represent the exact time when the images were taken. However, by visually comparing the data from different sensors, this delay was eliminated by applying approximated offset in the post process.

Capture software

Multithreaded C++ code was written to stream data from the sensors and save temporally aligned images from Lucid Vision sensors at 7 fps. Other sensors ran at highest frequency they could reach. Each sensor had one thread for reading data from the sensor and another thread for saving the data into the high-speed storage devices. The data was saved in raw binary format so that the streaming and saving software could operate at high fps. Ouster software was used to save LIDAR data as raw UDP packets (.pcap). Data from the IMU was recorded with a socketCAN package as raw CAN bus data. GNSS data was saved locally on a separate GNSS logger. Moreover, the captured data was distributed across multiple SSDs to improve writing speeds.

While these sensor threads were running, they were listening to commands from a custom user interface to save data (see Figure 11). The sensor threads were sending their status (e.g., running or exception) to the interface.

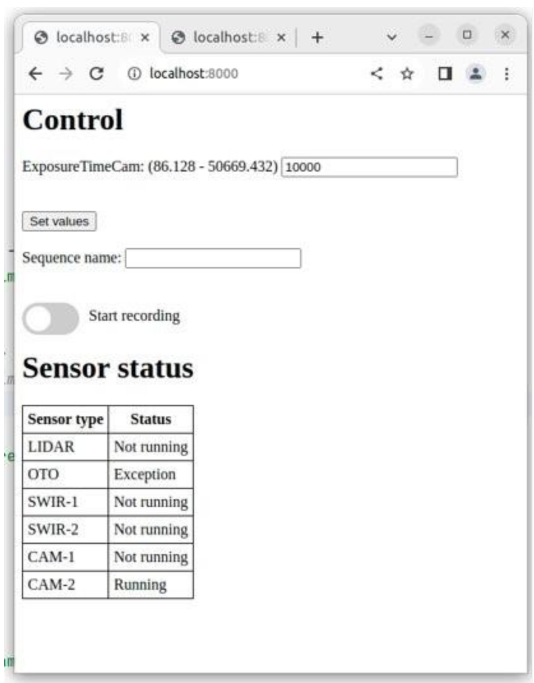


Figure 11: User Interface to record the sequences.

In addition, the user interface was used for defining the exposure of RGB cameras. The exposure setting was adjusted so that the RGB images were not over- or under-exposed at the start of the recording. If the environmental lighting changed drastically during one sequence, this value was adjusted in real-time. This setting value was constantly polled by RGB camera threads and adjusted in the nodemap in real-time. The user interface collected the status from the sensor threads. This would be useful for stopping the capture if one or more sensors stopped abruptly. The user could enter the sequence name and start and stop the recording using the toggle switch.

A separate user interface for the Senop camera was used to collect the hyperspectral data. Because the software took from 1 to 2 minutes to record one frame, the hyperspectral data was captured separately.

Dataset collection

The sequences were collected in a test environment while the wheel loader handled the objects of interest. The sequences were later post-processed to extract easy to use data from them. The data collection process and post-process are described in detail below.

Environment and work targets

The area behind the mobile lab (see Figure 12) at the Hervanta campus was used for collecting the dataset. The wheel loader was driven in front of selected work targets and the data from the sensors was collected. Different target objects such as containers and boxes (refer to Figure 13) were captured.

Figure 12: Test ground.

(a) (b) (c)

Figure 13: The captured work targets: containers (a), logs (b) and boxes (c).

The dataset was planned to be collected in both normal working conditions as well as in different challenging conditions during several days. Due to time constraints, the data collection was done on a single day. Table 2 lists the different factors in the dataset that could be varied. The factors which were captured are highlighted in bold. It was moderately sunny on the day of data collection.

Table 2: Different factors of dataset collection.

Factors	Types of options
Type of space	Indoor (cold storage hall for vehicles), **outdoor** (field outside the hall)
Lighting conditions	Different time of the day: **day** and evening Lights of the outdoor field on or **off** Lights of the vehicle on or **off**
Types of objects	Gravel, pellet, **boxes, container,** log, barrel, pit, tractor, safety cone
Weather conditions	**Sunny**, cloudy, rainy, snowy (depending on the timing of the data collection)
Vehicle speed	**Low speed, normal speed**
Different object distances	**Far away to close to pick object**

Post-processing of raw data

The collected data was post-processed. For easy usage, the data was converted to formats that are easy to read and access such as .png (portable network graphics), .csv (comma separated values) and .ply (polygon file format). Timestamp for each frame is either included in the filename or on the rows of a related .csv file. Timestamps in filenames are represented in UNIX format in milliseconds and in .csv files in seconds. The post-processing performed for the data from different sensors in summarized in Table 3.

Table 3: Post-processing performed for data from different sensors.

Sensor	Raw data	Post-processed data
Lucid Vision IMX490 RGB	Bayer pattern image	Debayered image (.png)
Lucid Vision IMX990 SWIR	Image	Format conversion (.png)
OToBrite oToCAM500 ToF	Depth (.png) and NIR (.png) images	Depth (.png) and Confidence (.png)
Ouster LIDAR OS0	Point clouds (.pcap, .json)	Point clouds (.csv, .ply)
Senop Rikola HSC-2	Data files (.hdr, .dat)	Images for each band (.png)
IMU	Raw CAN	.csv
NovAtel GNSS	Raw data	.csv. One for post-processed and one for matching FARO Laser scanner.
FARO Laser scanner	FARO scan	Point cloud (.ply)

Debayering function was applied to the raw Bayer pattern RGB images to convert them to 3-channel color images. No postprocess was done for the SWIR images. Depth images from the ToF sensor are provided as they come from the sensor. They are 16-bit images with 12-bits in use and where maximum value 7500 indicates 7.5 m distance. The confidence values for ToF camera were calculated from the NIR image according to the formula provided by the manufacturer. All images were exported as .png.

UDP packets from the Ouster sensor (.pcap) were converted with the Ouster studio to a more readable format. Point clouds for each frame were exported to .ply format.

Raw CAN bus data from the IMU were decoded with SavvyCAN software and exported as a single .csv file per sequence. Timestamps are included for each row.

The hyperspectral data from Senop was collected in 1000 bands in the range of 400 nm to 1000 nm and stored as .hdr and .data files. The data was stored as a compressed format. Each band information was extracted as 1000x1000 pixel image and stored as .png image.

FARO Laser scanner captures were post-processed, combined in FARO's software and exported in .ply format. Point cloud data was further processed by orienting it in such a way that positive y-axis points to the north and positive x-axis to the east.

GNSS data was processed with NovAtel's Inertial Explorer software. Additional base station data from Tampere University's antenna was used to apply differential GNSS correction to it. The resulting geographic data was exported into .csv format. Additionally, another file of this was created, where the values are scaled to match the point cloud from the FARO Laser scanner.

The exact location of the IMU sensor and GNSS antennas was read from the sensor rig sketch. The positions of these are reported in respect to the left RGB camera. Intrinsic and extrinsic calibration of the other sensors was done with MATLAB's camera calibration toolbox.

Dataset folder structure

The dataset can be downloaded at https://doi.org/10.23729/a39c02f5-e167-43e2-ac4e-815313619f20.

The collected data is arranged in folders, one folder for each sequence. Common data such as calibration matrices are stored outside of the folders. Two types of data are provided: sequence and snapshot. The snapshot has one frame for each sensor and one postprocessed frame from Senop. The sequences consist of five minutes of data, but they do not have Senop data.

- `GT.ply`
- `Parameters`
    - `RGB1_K.txt`
    - `RGB2_K.txt`
    - `SWIR_K.txt`
    - `ToF_K.txt`
    - `RGB1_distortionCoefficients.txt`
    - `RGB2_distortionCoefficients.txt`
    - `SWIR_distortionCoefficients.txt`
    - `ToF_distortionCoefficients.txt`
    - `RGB1_RGB2_T.txt`
    - `RGB1_SWIR_T.txt`
    - `RGB1_ToF_T.txt`
    - `RGB1_LIDAR_T.txt`
- `Sequence_1`
    - `GNSS_1.csv`
    - `GNSS_1_gt.csv`
    - `GNSS_2.csv`
    - `GNSS_2_gt.csv`
    - `IMU.csv`
    - `Lidar.json`
    - `Lidar.pcap`
    - `Lidar`
        - `Idx_timestamp.ply`
        - `...`
    - `Lidar_all_frames`

        - Idx_timestamp.ply
        - …
    - RGB1
        - Idx_timestamp.png
        - …
    - RGB2
        - Idx_timestamp.ply
        - …
    - SWIR
        - Idx_timestamp.png
        - …
    - Depth
        - Idx_timestamp.png
        - …
    - Depth_all_frames
        - Idx_timesamp.png
    - ToF_NIR
        - Idx_timestamp.png
        - …
    - ToF_NIR_all_frames
        - Idx_timestamp.png
        - …
    - Preview
        - Idx.png
- Target_object_1_snapshot
    - Senop
        - Band.png
    - LIDAR
        - 0.ply
        - 0.csv
    - Color1
        - 0.png
    - Color2
        - 0.png
    - SWIR
        - 0.png
    - TOF
        - nir.png
        - depth.png

The file `GT.ply` is the processed ground truth point cloud from the FARO Laser scanner.

`GNSS_1.csv` contains the geographic data from the antenna on the front and `GNSS_2.csv` from the back of the vehicle. These values are further scaled and adjusted to create `GNSS_1_gt.csv` and `GNSS_2_gt.csv` where X and Y positions refer to the positions in the recorded ground truth point cloud.

`IMU.csv` contains extracted data from the IMU sensor. Timestamps in these files are in UNIX format in seconds.

Folders `RGB1`, `RGB2` and `SWIR` contain temporally aligned data where same indexes match with each other. Folders `Lidar`, `Depth` and `ToF_NIR` contain closest frames from these sensors corresponding to `RGB1` frames. Folders with extension "`_all_frames`" also contain all in between frames for the LIDAR and ToF sensors. Indexing in these does not match with other sources and timestamps should be used instead. Timestamps are in UNIX format in milliseconds.

Original raw data from the Ouster LIDAR sensor are provided in .pcap and .json formats along with extracted point clouds. These can be opened, for example, with Ouster Studio software or with Ouster SDK. Other formats such as .png can be opened in OpenCV (https://opencv.org/) and MATLAB whereas .csv and .ply can be opened in MATLAB or with libraries such as Open3D (http://www.open3d.org/) and PCL (https://pointclouds.org/).

Intrinsic matrices and distortion coefficients for each sensor are in the text files under parameters folder. Poses of other sensors are reported as transformation matrices in RGB1 camera's coordinate frame as illustrated in the Figure 14. These can be directly utilized to transform points from these sensors to RGB1 camera's coordinate frame.

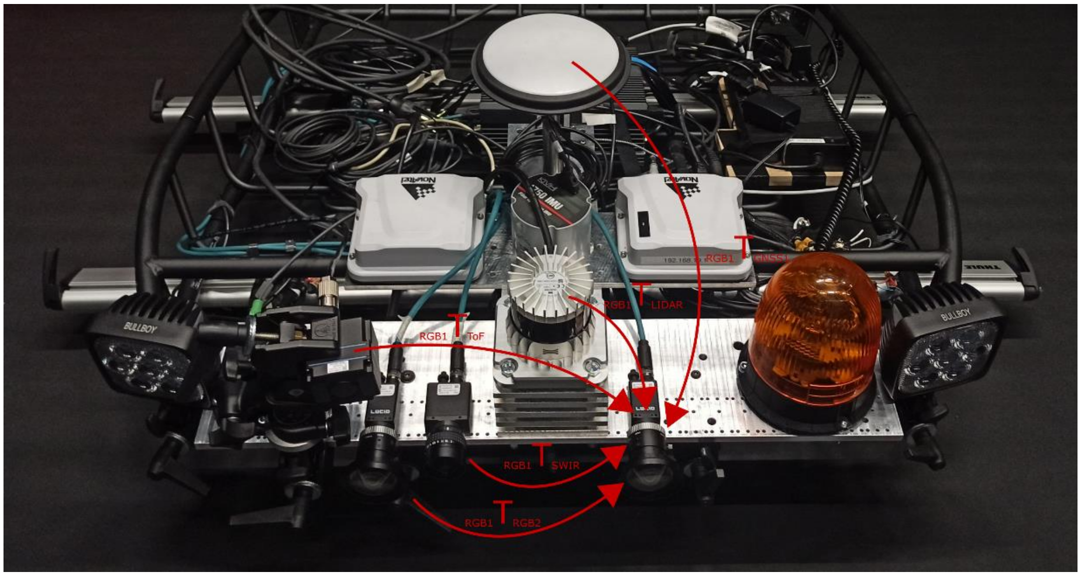


Figure 14: Transformations.

Sample data

The ground truth 3D data from the FARO Laser scanner is shown in Figure 15.

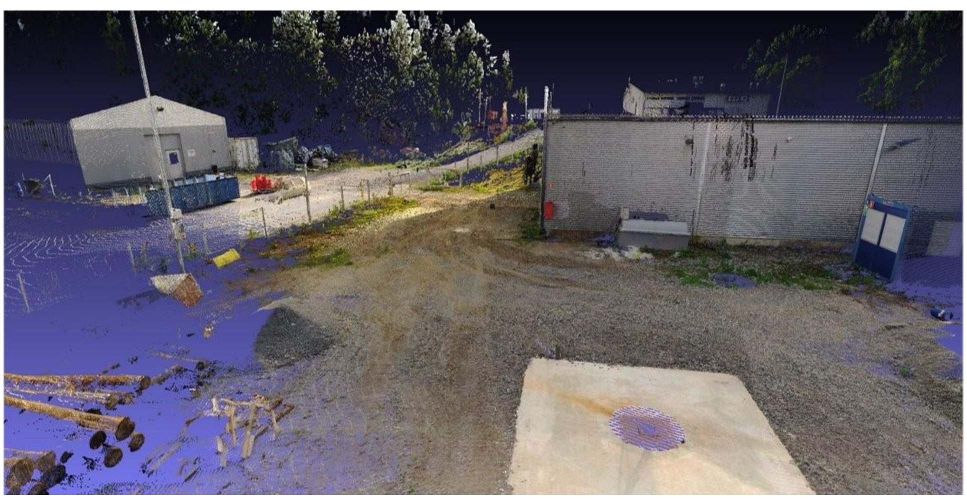
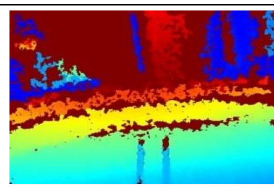
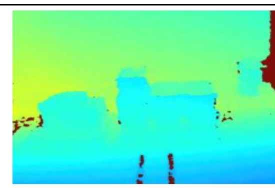
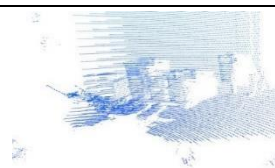


Figure 15: Processed data from the FARO Laser scanner.

Sample images and point cloud data for different target objects are shown in Table 4 to illustrate differences in the visual data between different sensors.

Table 4: Sample data from some of the sensors.

Sensor type	Container	Box
Lucid Vision IMX490 RGB		
Lucid Vision IMX990 SWIR		
OToBrite oToCAM500 ToF		
Ouster LIDAR OS0		

Conclusion

The collected dataset is useful to explore 3D multimodal sensors for work target detection and measurement applications for work machine industry. This dataset helps to overcome certain limitations of color cameras. Using information from color cameras is challenging when it comes to dark lighting conditions, changing light directions, weather conditions and occlusion. Depth sensors such as LIDAR and ToF will work in dark environments and will be insensitive to changes in illumination and sun direction changes. However, depth sensors will sense objects that appear first in the scene and hence occluded objects cannot be sensed through depth sensors. Hyperspectral cameras provide information related to the materials, and they are insensitive to illumination, occlusion and work in different weather conditions. Using a combination of different modalities helps to overcome these challenging conditions.

By carrying out this PoC tailored for the FIMA companies, we were able to share information on CIVIT's imaging equipment and expertise as well as to develop the capability to use the equipment for collecting custom datasets in a quick and cost-efficient way. The shared open dataset is an example illustrating the possibilities of multimodal 3D imaging. In the future, it is possible to collect more extensive and customized datasets based on the needs of interested companies.

References

1. MATLAB code used for camera and LIDAR calibration
   https://www.mathworks.com/help/lidar/ug/lidar-and-camera-calibration.html
2. MATLAB code used for camera calibration
   https://www.mathworks.com/help/vision/camera-calibration.html