Shared Dataset README

This folder contains two sample datasets: Daytime and Nighttime. The Daytime dataset will contain 5784 RGB frames and their corresponding 5784 of Infrared frames. The Nighttime dataset will contain 8595 infrared frames only, as tabulated in Table 1.

Dataset	No of Frames
Daytime	5784 RGB frames and 5784 Infrared frames
Nighttime	8595 infrared frames only

Table 1: Sample Dataset

Several state-of-the-art sensors have been used to build the dataset, such as LiDAR, infrared camera, stereo camera, IMU, and GNSS. All the sensors are mounted on a single Unmanned Ground Vehicle (UGV) with a maximum speed of 1.5 m/s. The data from all these sensors have been recorded using ROS (Robot Operating System) Tool called rosbag. The information about all the data is tabulated in Table 2:

Data	FPS (Frames Per Second)	Sensor Used
Infrared Camera Raw Image	15	Teledyne FLIR BOSON
Infrared Camera Rectified Image	15	Teledyne FLIR BOSON
Stereo Left Rectified Image	15	StereoLabs ZED2i
Stereo Right Rectified Image	15	StereoLabs ZED2i
3D point cloud	15	Ouster OS0-64
IMU	200	Vectornav VN300
GNSS	8	Ublox F9P

Table 2: Supporting Data in the ROSBAG.

From the rosbag, the rectified infrared frames and the rectified stereo left RGB frames (that have been transformed using the homography method to match the infrared image) are extracted. Each extracted frame has been saved along with its timestamps: rosbag and header, and then organized in its respective folder with a JSON file. The JSON file contains ground truth information like the boundary box position, occlusion, and truncation level. More details regarding the JSON file are tabulated in Table 3.

Attributes	Explanation
Boundary Box	The 2D position of the boundary box in the image frame
Occlusion	Level 0: No occlusion at all
	Level 1: Occlusion of the object less than 50%
	Level 2: Occlusion of the object more than 50%
Truncation	Level 0: No truncation at all
	Level 1: Truncation of the object less than 50%
	Level 2: Truncation of the object more than 50%

Table 3: JSON information

The other data such as 3D Pointcloud, IMU, GNSS are extra supporting data, and they are OPTIONAL for the participants to use for their solution in this challenge.
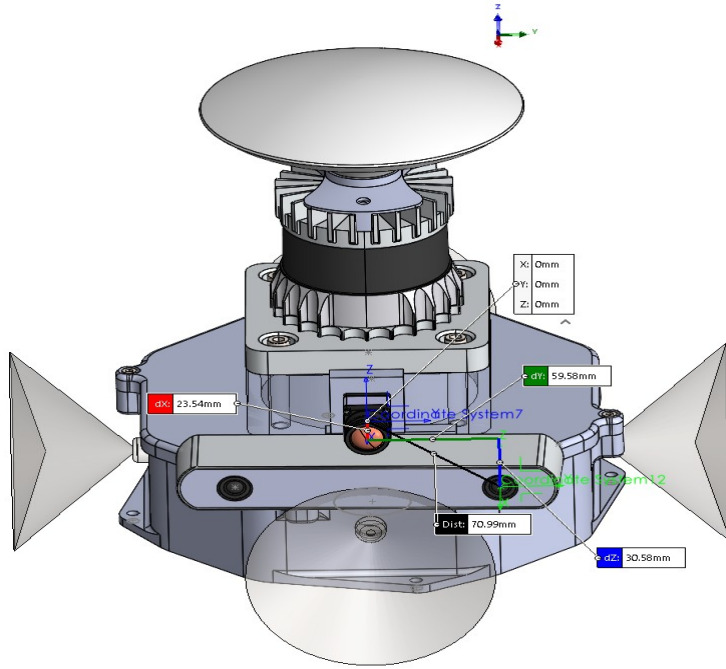
The extrinsic of all the sensor's positions with respect to the infrared camera are tabulated in Table 4:

Note that all sensors share the same frame (right-hand-rule) in which the axis X+ is facing forward.

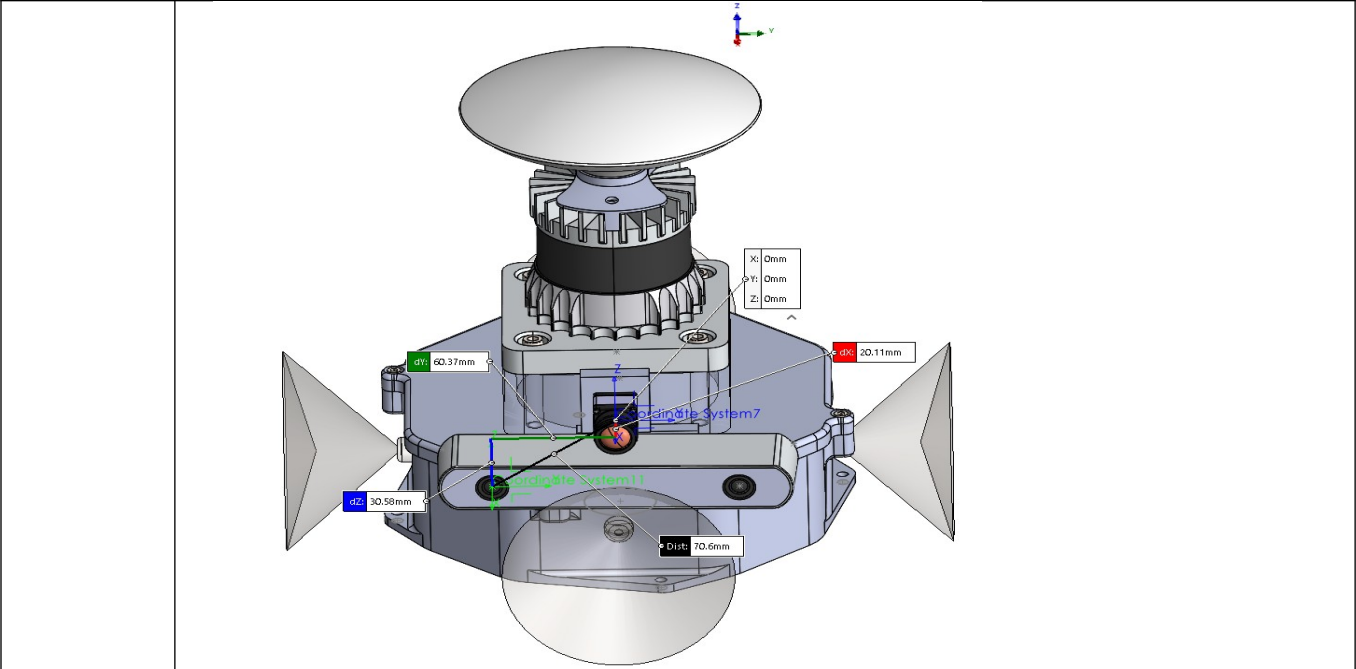| Sensors	Extrinsic [X,Y,Z,ROLL,PITCH,YAW] [mm,mm,mm,deg,deg,deg]
FLIR Boson Infrared Camera Raw Image	[0,0,0,0,0,0] Reference Point

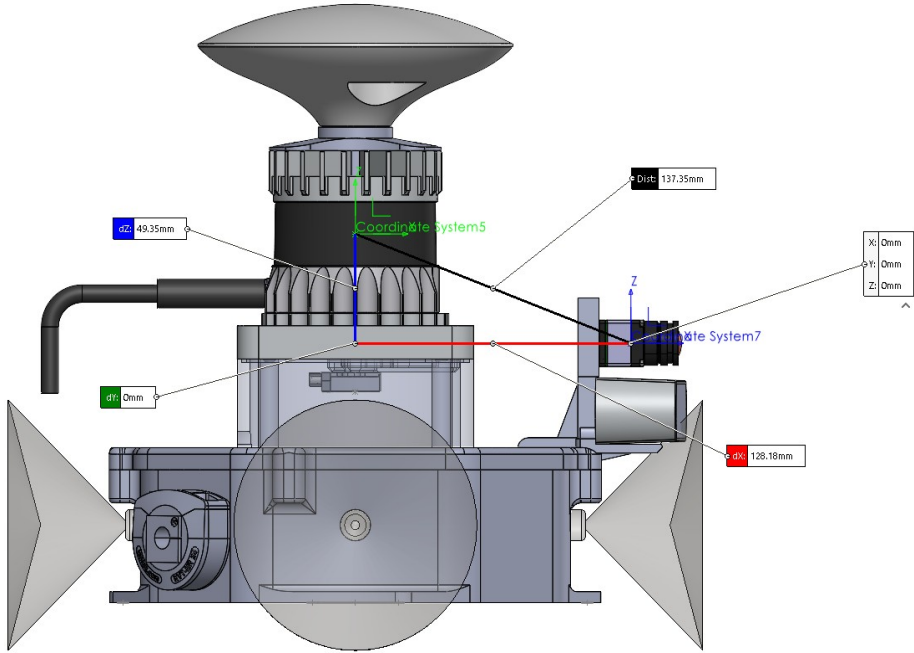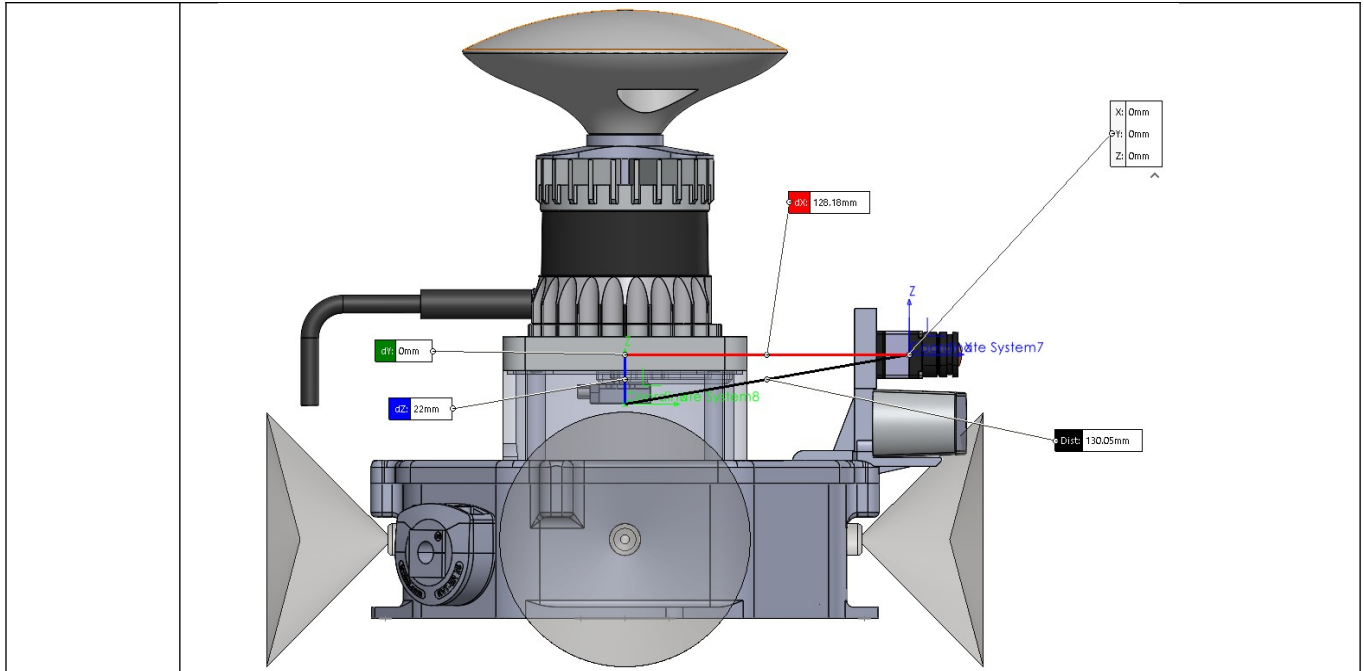|	
ZED Camera Stereo Left Rectified Image	[23.54, 59.58, -30.58, 0, 0, 0]
ZED Camera Stereo Right Rectified Image	[20.11, 60.37, -30.58, 0, 0, 0]

| Ouster-OS0 3D Pointcloud | [128.18, 0, 49.35, 0, 0, 0] |

| Vectornav IMU | [128.18, 0, -22.0, 0, 0, 0] |

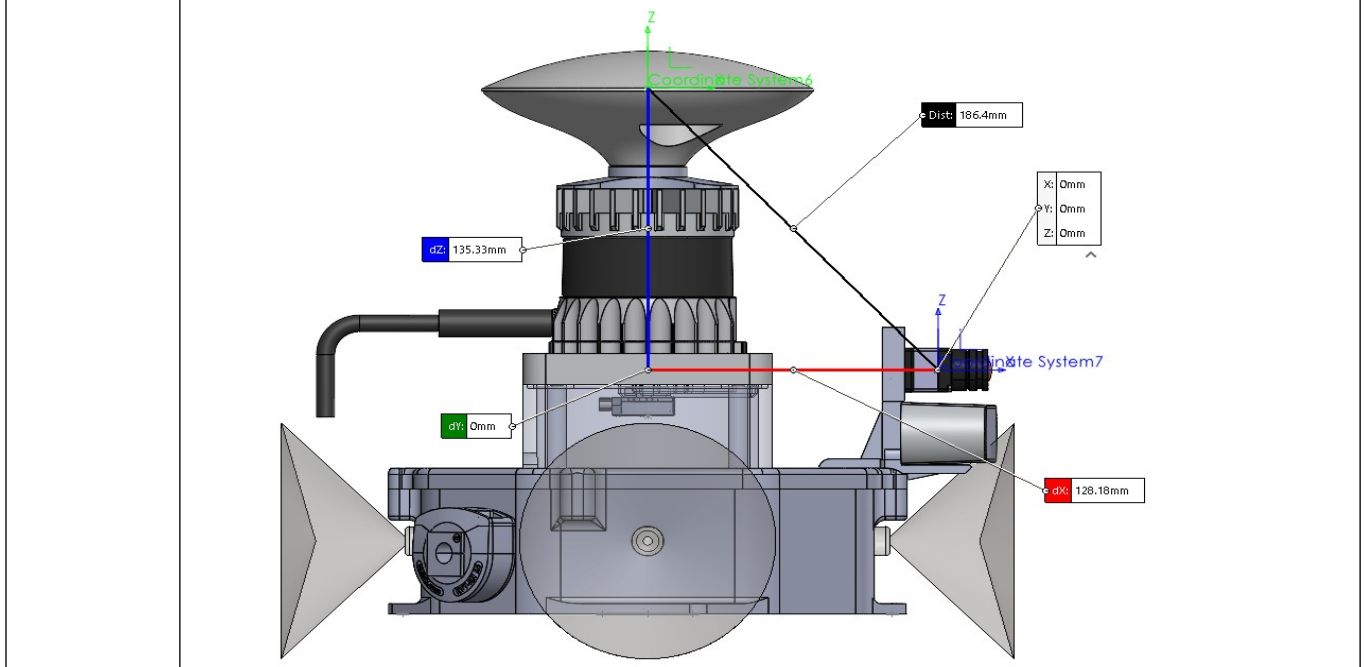|	
GNSS	[128.18, 0, 135.33, 0, 0, 0]

Table 4: Sensor's Extrinsic Information with Respect to the FLIR Infrared Camera

The camera matrix and distortion coefficients for each camera are tabulated in Table 5.

Sensors	Intrinsics camera matrix (3x3), K and Distortion coefficient, D
FLIR Boson Infrared Camera Raw Image	K: [408.3403051749102, 0.0, 407.62152914104007, 0.0, 316.7098391031272, 253.81677458989378, 0.0, 0.0, 1.0] D: [-0.3726371743741999, 0.1357631218268681, -0.0007926975585932571, -0.000391058346169 5004, 0.0]
ZED Camera Stereo Left Rectified Image	K: [531.6238403320312, 0.0, 624.031982421875, 0.0, 531.6238403320312, 376.97601318359375, 0.0, 0.0, 1.0] D: None
ZED Camera Stereo Right Rectified Image	K: [531.6238403320312, 0.0, 624.031982421875, 0.0, 531.6238403320312, 376.97601318359375, 0.0, 0.0, 1.0] D: None

Table 5: Camera Matrix and Distortion Coefficients for ZED Stereo and FLIR Infrared Cameras

The folder and files are organized as below:

- **README.pdf**

- **Daytime**   **(folder)**
  - o **RGB**   **(ZIP file)**
    - ▪ **imageXXXX.png**
  - o **IR**   **(ZIP file)**
    - ▪ **imageXXXX.png**
  - o **rosbag**   **(folder)**
    - ▪ **rosbag.bag**
  - o **daytime.json**   **(annotations file)**
  - o **rgb_timestamps.csv**   **(synchronization file)**
  - o **ir_timestamps.csv**   **(synchronization file)**

- **Nighttime**   **(folder)**
  - o **IR**   **(ZIP file)**
    - ▪ **imageXXXX.png**
  - o **rosbag**   **(folder)**
    - ▪ **rosbag.bag**
  - o **nighttime.json**   **(annotations file)**
  - o **ir_timestamps.csv**   **(synchronization file)**

- **sample_result.json**   **(expected output format / sample)**

Note for .csv files:
These files are used to synchronize the zipped images with data from other sensors, in case the solver wants to use data from either one / all other sensors from .bag

file. Point of synchronization is the first column, while the second represents the timestamps from sensors' drivers and is given only as a reference.

Notes on sample_result.json:
1. This file specifies the expected output format of an object tracker
2. Concretely, this is an output of a tracker that tracks 3 objects (with ids 3, 8 and 11) in 5 frames (with ids 1, 2, 3, 4 and 5)
3. "predictions" is an ordered array of tracker's predictions. Each prediction has its own id, an image id in which the object is located, the id of tracked object, object's location in bbox coordinates, and a confidence score of prediction
4. "images" is an ordered array of images in which predictions were made. If there is no single prediction in an image – that image must not be in this array. Each image has it's id and file name.
5. "objects_tracked" is an ordered array of objects being tracked. Each object has it's id. In case the tracking solution also does classification, then the class label should be put here as well.
6. "predictions", "images" and "objects_tracked" all have an additional field "extra_dict": this is a field left to solvers to fill in case their solution outputs more features and they wish to communicate this with us. Keys of this dictionary would be the added features and values would be the details of each added feature.
7. Please update "info" accordingly.