Homework Project

The main sensors used at Kiwibot are cameras because they are cheap and give us very rich and dense information. The main problem with cameras is their processing, which in most cases is non-trivial. Here is where DL comes to an important play.

Our latest model of the robot has 5 cameras surrounding the robot, one on each side and 2 on the front. We want to reconstruct the environment around the robot. To do that we want to map the 2D images to 3D space, which in this case is called monocular depth estimation.

We built a dataset consisting of 4 cameras (just one on the front), GPS, and a 3D lidar mounted on the robot. The dataset given includes at each timestamp, some metadata, the 4 images from the cameras, and its corresponding aligned depth map coming from the 3D lidar.

The structure of the dataset is:

```
📦dataset
 ├ 📜camera_color_image_raw_depth_map_0.png
 ├ 📜camera_color_image_raw_image_0.jpg
 ├ 📜metadata.csv
 ├ 📜video_mapping_back_depth_map_0.png
 ├ 📜video_mapping_back_image_0.jpg
 ├ 📜video_mapping_left_depth_map_0.png
 ├ 📜video_mapping_left_image_0.jpg
 ├ 📜video_mapping_left_image_1.jpg
 ├ 📜video_mapping_right_depth_map_0.png
 └ 📜video_mapping_right_image_0.jpg
 └ 📜 ...
```

The CSV file has all the metadata associated with the images and the name of the corresponding image.

The CSV has the following columns:

- */video_mapping/left:* name of the image file of the camera on the left of the robot
- */video_mapping/left/depth_map:* name of the depth map file of the left camera
- */video_mapping/right:* name of the image file of the camera on the right of the robot
- */video_mapping/right/depth_map:* name of the depth map file of the right camera
- */video_mapping/back:* name of the image file of the camera on the back of the robot
- */video_mapping/back/depth_map:* name of the depth map file of the back camera
- */camera/color/image_raw:* name of the image file of the camera on the center
- */camera/color/image_raw/depth_map:* name of the depth map file of the center camera
- *lat*: latitude
- *lon*: longitude
- *timestamp*: Unix timestamp in milliseconds associated with cameras

- All other columns are self-explanatory. Units of angles are radians and velocities are m/s. The ***x-axis*** is going forward to the robot, and the ***y-axis*** is to the left.

Note on image formats: The camera images are just normal JPG files. The depth maps are PNG *uint16* images, where each pixel corresponds to depth in millimeters.

Some comments:

- Please create a private repository on GitHub and share it with us once you finish. The repo should be self-contained and it should have a README with instructions on how to run the code.
- Please avoid uploading to GitHub any data from the dataset.
- You are free to use any DL framework, but we recommend either TensorFlow or PyTorch. You can use any high-level API (Keras, Pytorch-lightning, etc).
- You can use Google Colab to train the model(s) since it is free :) An alternative is Kaggle Notebooks which is also free but can run code for longer without interruptions.
- We won't be evaluating the model performance (that much), but how you approach the problem, how you structure the solution, and how you present it.

The tasks are:

1. Build a model for depth estimation using all 4 cameras. The model should process the 4 cameras at the time and output the corresponding 4 depth maps
2. Prepare a basic presentation on how you approached the problem and show some results.
3. [Extra]: Build a model like in (1), but now include time in the model (take into account a sequence of consecutive images at a time)

Dataset

The dataset is compressed in this file. If you need access please request it directly with your email.

A snapshot of the dataset can be seen in this video.