• What did you learn after looking on our dataset?

Images have been taken from a fisheye camera lens with FOV > 90°.

Image sizes are not the same.

Some files were not having any image content.

Most of the images look very similar at first glance.

• How does you program work?

1. Resize all images to a single image size – (640, 480)
2. Run a nested loop to gather the previous and the next image
3. Comparison of the frames based on different image processing techniques.
4. A higher score means less similarity.
5. So any score value lesser than a particular threshold are meant to have similar images and therefore noted to be removed later.

• What values did you decide to use for input parameters and how did

you find these values?

1. One kernel size for the Gaussian blur was decided based on what causes similar images to be removed more. A kernel size of (5,5) was better at observing dissimilarity than (3,3).
2. Threshold values – Lower threshold : 127, higher threshold: 255
3. Minimum Contour area was decided to be 10

• What you would suggest to implement to improve data collection of

unique cases in future?

1. Maybe using similar size images can help to compare the images better.
2. Detection of similar images shouldn't be done in a single run. First we detect absolutely different small set of images and then for any image captured by the camera, verify these for similarity with the already present images.

• Any other comments about your solution?