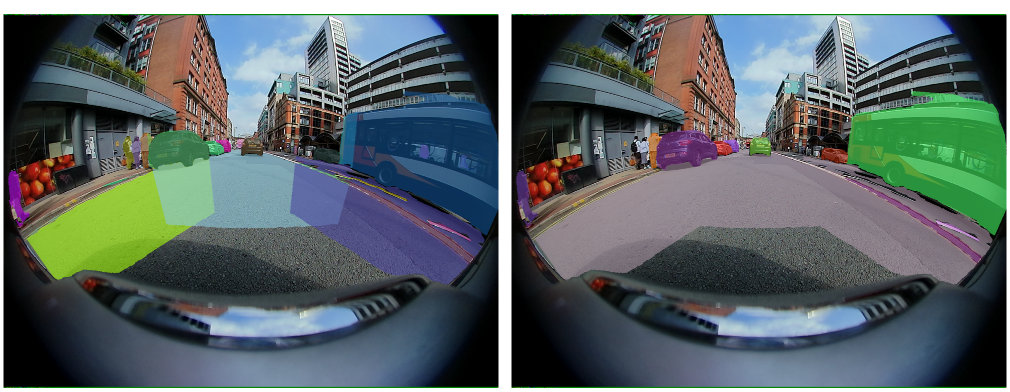
Object Detection and Segmentation using Fisheye Images

Theodora Gaiceanu and Felix Slothower

Computer vision applications are everywhere, one of their purposes being to make the world safer, whether it is about an airport, a crowded city, or traffic. A way to make traffic safer by decreasing the number of accidents is to use autonomous cars. However, self-driving vehicles need to meet a long list of requirements in order to be allowed on the roads, one of these being to have a full panorama view of their environment. This is where fisheye cameras could come to the rescue!

A fisheye camera uses a specific lens that has a wide field of view, capable of seeing objects entirely to the left and right of the direction the camera is pointing. Having a field of view of roughly 180 degrees, a full coverage of the surroundings of the vehicle can be ensured with no more than four cameras, one on each side of the car. While this seems a cheap and straightforward solution, one needs to be aware that using this kind of image introduces some challenges. A key step in the pipeline for autonomous vehicles is to classify and highlight relevant objects surrounding the vehicle. These highlighted regions are referred to as segmentations. One challenge when using fisheye images is that the classical models used to identify and localize objects in an image can have a hard time with this kind of images due to their different appearance. While the objects located in the center of the image may seem normal, they can be very deformed at the corner of the image (e.g notice that the car in the middle of Figure 1a looks normal, but the bus is bent).

One solution is to alter the fisheye image digitally, after it has been taken, so that it resembles a more standard image, a process called rectification. With this in mind, we tried two rectification techniques and evaluated the results. The first rectification technique had the advantage of maintaining the wide field of view and was able to straighten objects vertically. However, the results that we got using this technique were worse than the results with the fisheye images. The second technique resulted in a significant loss of information, since it was cropping the image in order to keep a consistent scale with respect to objects at the center of the image. This correction technique is called rectilinear correction. Using this method did provide improved results when only considering the objects remaining in the cropped image, but a different approach would be needed if the original field of view was going to be maintained. Consequently, another approach was needed. The last method that we tried was to divide the fisheye image into 3 or 6 tiles and to use the rectilinear correction on each tile separately. Afterward, a merging of the segmentations from each tile was performed in order to produce one unified set of segmentations for a given image. This process is illustrated in Figure 1, with unmerged segmentations on the left and merged segmentations on the right. This method had the advantage of keeping the wide field of view while removing the warping of objects during the prediction stage, and achieved the best results.



(a) Unmerged segmentations **(b)** Merged segmentations

Figure 1