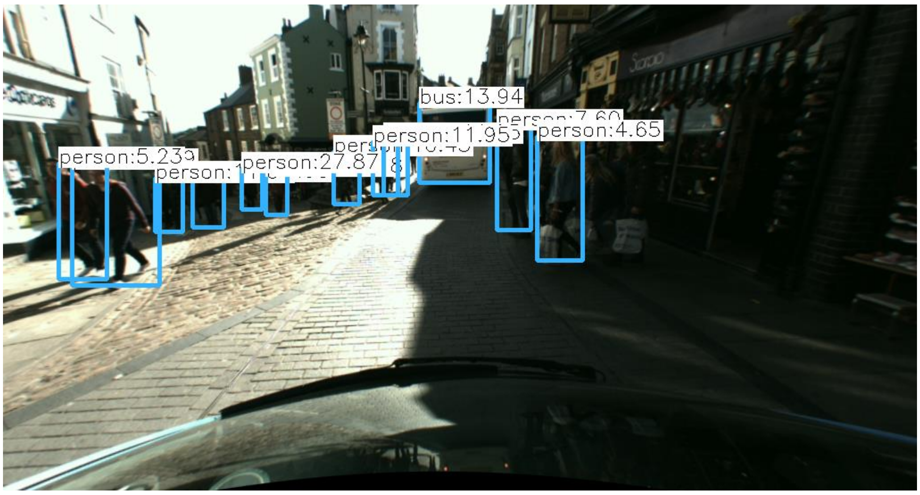
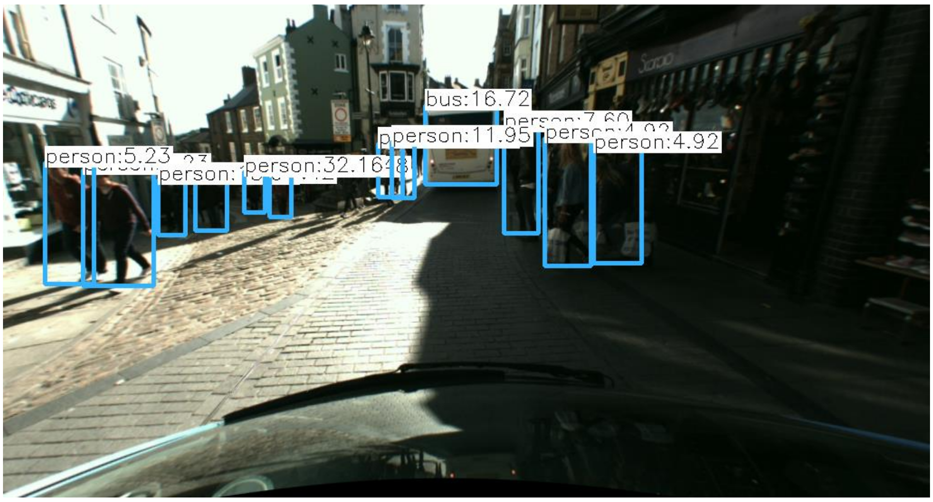
Computer Vision Report

A lot of images in the dataset have large areas of low contrast, such as in shadows or in areas of glare. These will reduce the performance of any object detection or distance estimation. A common technique to improve the contrast in a scene is histogram equalization but if an image has a large dynamic range this technique can underperform, for this reason I apply Contrast Limited Adaptive Histogram Equalization (CLAHE) [1] which uses multiple histograms over the image so improve the contrast.



1506943031.476865_L Feature Detection - No CLAHE

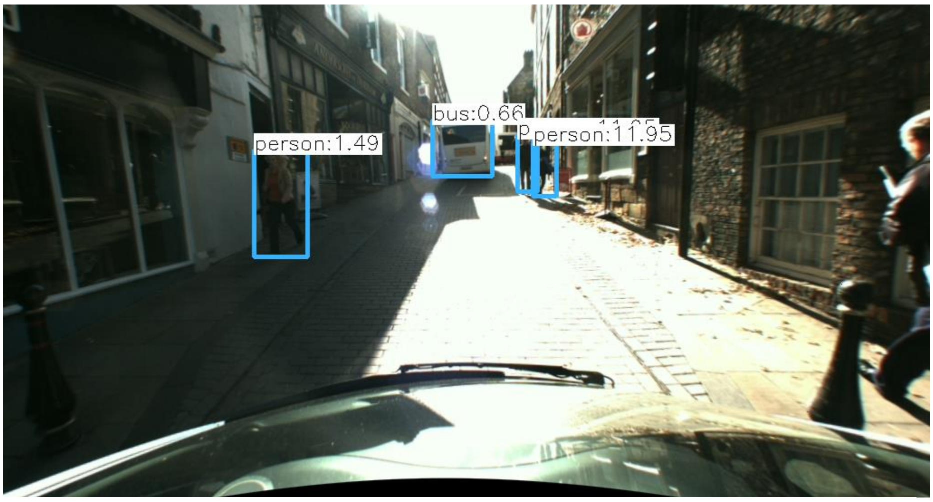


1506943031_476865_L Feature Detection – With CLAHE

The image with CLAHE applied identified a person originally hidden in the shadows, however it lost a person in the middle left which was identified by the original. One can consult Figure 1 in the Appendix to observe the contrast increase in the dark regions to see how the extra person was identified. Whilst we did lose a dynamic object, the nature of the system means we prioritise objects nearest the cameras as they have a greater influence on the decisions made by the vehicle such as whether it should speed up or brake meaning CLAHE is advantageous to our system.

To improve the disparity map, we use a Laplacian filter to remove noise then a powerwise operator ($x^{0.75}$) to the greyscale left and right stereo images.

During my design I investigated whether a sparse stereo approach could be more efficient then the dense approach. Using ORB, varying the number of feature points between 1000 and 6000, and a brute-force matcher I found that often a sparse approach can recover inadequate or incorrect distances for dynamic objects.



1506943061.468682_L.png using sparse stereo with ORB (6000 feature points)

Clearly the leftmost objects have incorrect distance information. My hypothesis was not enough feature points were getting matched for objects so only applied the ORB matcher to the bounding boxes however in a lot of images ORB failed to find feature points, I tried expanding the bounding boxes to see if I was cutting features off still the sparse approach was proving ineffective.

I applied the Joint Bilateral Filter ($\sigma_{color} = 3$, $\sigma_{space} = 8$) to the disparity map to reduce any noise from it using a greyscale left image in for the disparity kernel. The resultant images can be seen in Figure 2, 3. If one were to zoom in, you'd see there is less noise in the objects but the computation time required compared to the negligible performance increase renders the JBF not useful.

I investigated multiple options when recovering the distance from the dense disparity map. Let D be all the values of the pixels in the disparity map for some object. My first approach was to take the mean of D however often D contained a lot of zero-values meaning the distance was always an overestimate. My second approach was to take the mean of all the non-zero values of D but this method was particularly sensitive to noise. My final approach was to take the 80% percentile point of the non-zero values of D which proved to be robust to noise.

To quantitively evaluate the performance of my system I selected several photos from the dataset that are particularly challenging for the system to annotate and compare the number of objects detected to the actual number in them.

Image Name	Image Details	#Dynamic Objects Labelled	#Dynamic Objects in Scene
1506942568.478817_L.png	Large area of glare and shadows	2	4
1506942578.476638_L.png	Large number of objects in	8	8

	dark and bright regions		
1506942792.477428_L.png	Objects hidden in very dark region	5	8
1506943031.476865_L.png	Large number of objects close to each other	11	14
1506942817.476407_L.png	Large number of people	11	14
1506943010.480501_L.png	Objects at difference distance from camera	14	14

The systems performance was independent of the number of objects, but naturally a larger of objects did take longer to process. The system performed well against changing conditions always being able to identify the close dynamic objects in the scene, the large areas of dark and bright regions always had a large proportion of the present objects labelled. The system did not cope well with objects that were far away from the camera, often they were identified but a distance could not be recovered, or the distance recovered was incorrect. Sometimes dynamic objects were clumped and were falsely identified as one object which prove problematic if one was performing tracking however for the current specification this is not seen as a negative trait. The speed of the video provided is 3 frames per second.

References

1. Pizer, S., Johnston, R., Ericksen, J., Yankaskas, B. and Muller, K. (1990). Contrast-limited adaptive histogram equalization: speed and effectiveness. *[1990] Proceedings of the First Conference on Visualization in Biomedical Computing*.

Appendix

Figure 1 1056943031.475865_L.png with CLAHE

Figure 2 1056943031.475865_L.png disparity map with no JBF

Figure 3 1056943031.475865_L.png disparity map with JBF