Computer Vision - Stereo Vision for Object Distance Ranging - hqfl21

1 Outline

My implementation runs as follow: Pre-process the left and right images by denoising the images with a median blur and using CLAHE (Contrast Limited Adaptive Histogram Equalisation) to widen the range of available pixels without blowing out some areas of the image; Calculate the disparity map, using SGMB, and apply a WLS (weighted least squares) filter to the disparity map to reduce the number of holes, which would otherwise act as noise, without removing any of the critical information; Apply Yolo on the pre-processed colour image to find objects in the frame, filter to find relevant objects; Calculate the distance of each pixel using the disparity map and calculate the distance of each object and output this information. SGBM and Yolo implementations are slightly modified versions of the provided code.

Each frame takes an average of 1200ms to process on a 4-Core Intel i5 8250u @ 3.2GHz, with yolo and disparity calculations taking up the bulk of this time.

2 Pre-Processing

The median blur worked best with a kernel of size 5, increasing object detection by around 8-10% on a series of 4, 10-15 frame snippets, with negligible effect on the distance estimates (Figures 1a-c). It also helped to reduce the number of false detections (Figures 2a-b). I tested this against a bilateral filter and the median blur gave better results overall.

Figure 1a: Reference Frame

Figure 1b: No blur, vehicle (truck) is not detected

Figure 1c: Blur, vehicle (truck) is detected

Figure 2a: No blur, incorrect object is detected

Figure 2b: With blur, object is detected

I run CLAHE on the light channel of the images converted to the LAB colour space. This helps in scenes where the sun shines on half of the objects (Figure 3a), making the other half much darker. The woman detected on the right is further away, but the non-equalized image produces a disparity map which estimates the woman to be closer (Figures 3b-e). The improvement can be seen in the disparity map (note: all disparity maps shown in this report have been histogram equalized to make them much easier to read. This means that the values in the disparity maps aren't comparable with other maps, but judgements can still be made).

Figure 3a: Reference image

Figure 3b: Without CLAHE, distance is predicted incorrectly

Figure 3c: With CLAHE, distance is predicted correctly

Figure 3d: Disparity map without CLAHE, values on the body are not the closest (brightest)

Figure 3e: Disparity with CLAHE, closest (brightest) values are on the body

I could not find any evidence, both in research and testing, to suggest that pre-processing the image using CLAHE improve object detection.

These disparity maps have already been passed through the WLS filter, hence very few holes and a low amount of noise. An example of the improvement WLS does can be seen in Figures (5a-b). Sometimes WLS can give a worse depth estimate, this is most likely due to it losing critical information when smoothing.

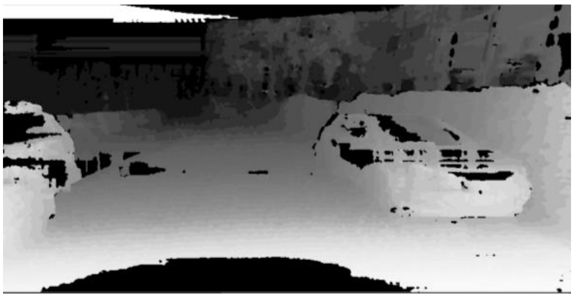

Figure 5a: Disparity map without WLS

Figure 5a: Disparity Map with WLS

3 Getting Depth

Once the bounding boxes are found, I found that taking the median of the distance values situated between the 20% and 50% quartiles (where 0 is the closest) to be the best (Figures 6a-c).

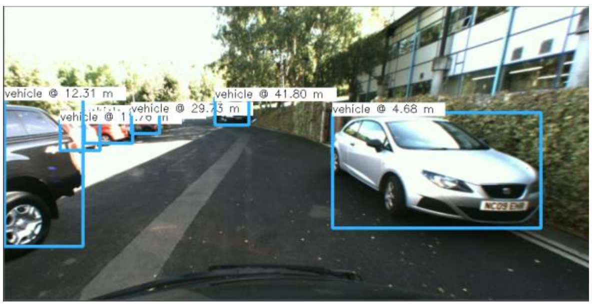

Figure 6a: Using the mean value, incorrect value on the left

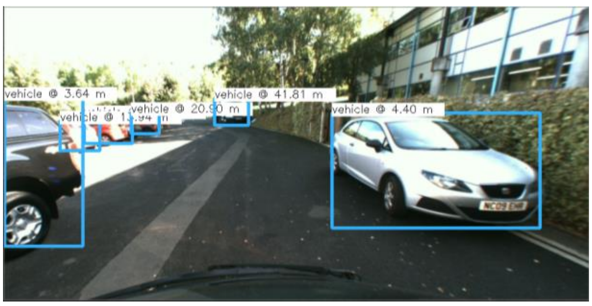

Figure 6b: Using the median value

Figure 6c: Using the mean of the values sliced according to percentile

The mean is very easily skewed by outliers (Figure 6a) and is therefore mostly useless. As the bounding box contains other distances which aren't the object itself, the median can sometimes be taken from an incorrect position. Taking the mean values from within the 20-50% percentile ensures both that outliers are removed and only distances within the object (namely, the closest ones) are considered.

4 Heuristic

I have tried to implement a system which changes which region the mean is taken from when a different type of object is detected. For a person, I take the between 50% and 80% of their height and between 30% and 50% of their width, to try and find their torso – the area with the least amount of background. This saw an overall improvement, e.g. in Figures 7a-c. For a vehicle, I try to select the bottom half, as this avoids the reflection/transparency of the windows, but not all the way down as when a vehicle is at a diagonal, there's more ground in the box. This saw no overall improvements, varying between an increase and decrease of performance.

Figure 7a: Reference Image

Figure 7b: Distance without a heuristic, appears to overshoot the estimate

Figure 7c: Distance with heuristic, guess is more accurate