Computer Vision Coursework

There are four stages in the computer vision pipeline developed for this project: image pre-processing, disparity map calculation, object detection, and object distance calculation.

**Pre-processing**

The main purpose of pre-processing is to remove photometric distortion. This can take many forms, but the main distortions in the TTBB dataset are specular surfaces (particularly wet roads), transparent objects (such as windows), greatly varying brightness, occlusion, and road markings. Typical operations for reducing specular distortions include:

* Laplacian of Gaussian (LoG) filtering (Kanade et al., 1995)
* Subtraction of mean values computed in nearby pixels
* Bilateral filtering (Ansar et al., 2004)

The steps taken in pre-processing were based off those used in Deepa & Jyothi (2017), namely the use of a median filter, Weiner filter, and histogram equalisation. A median filter was used to remove salt-and-pepper noise. The image on the right has been median filtered with a small neighbourhood size.



Two types of histogram equalisation were tested: “standard” and contrast-limiting adaptive histogram equalisation (CLAHE). The purpose of histogram equalisation is to increase the contrast of images. Regular equalisation performs poorly when the image contains regions that are significantly lighter or darker than most of the image, as the contrast in those regions will not be sufficiently enhanced.

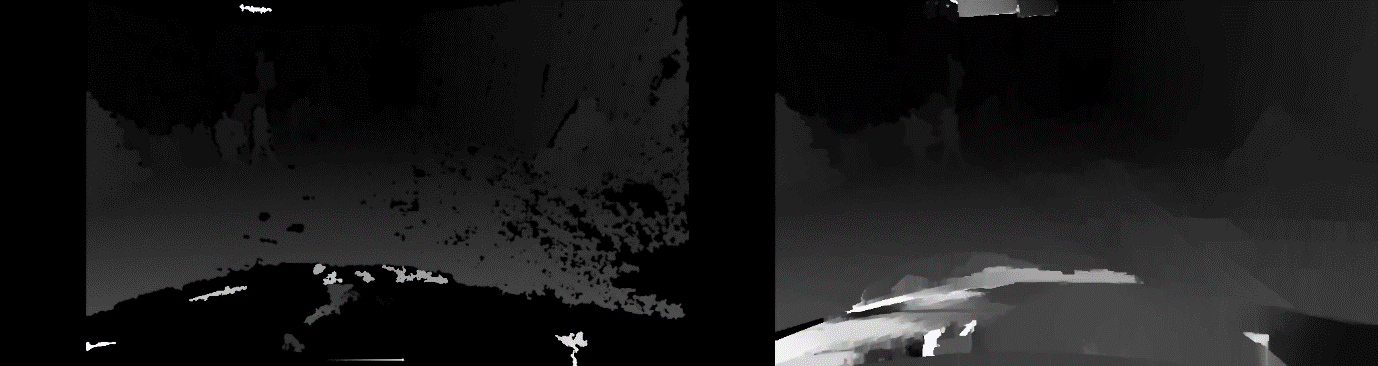
Adaptive histogram equalisation (AHE) combats this by computing multiple histograms, each corresponding to a distinct section of the image, and using them to re-distribute the lightness values of an image. AHE has a tendency to over-amplify noise in homogeneous regions, a shortcoming tackled by contrast-limiting AHE (CLAHE) first introduced by Zuiderveld (1994). The image below shows the original image, then the “standard” histogram equalisation, and then CLAHE.



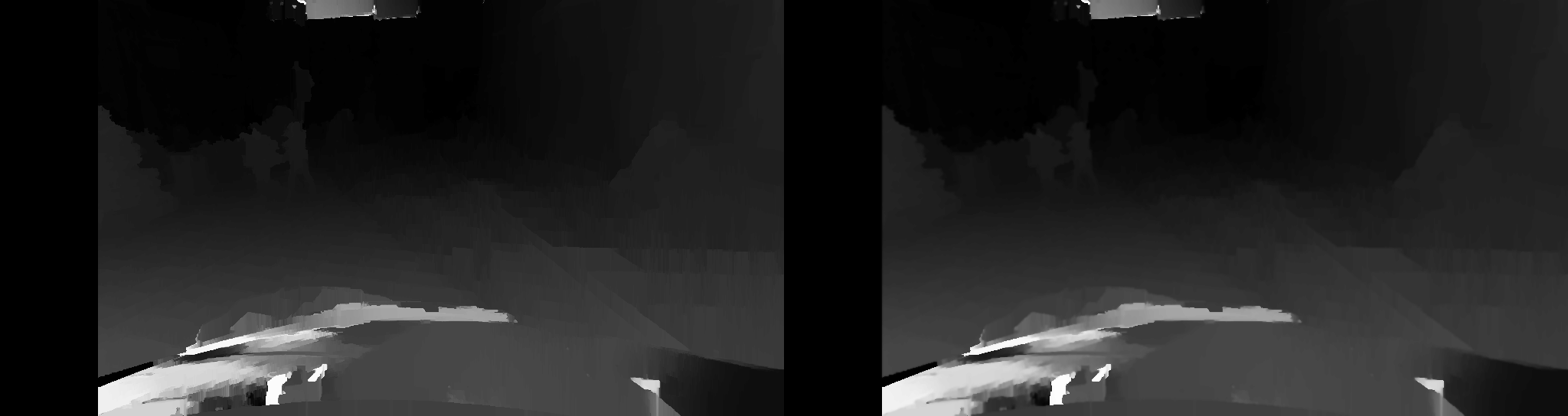
Details are a lot sharper, and details in the upper left-hand region have been preserved, while the lower right-hand region has been simultaneously lightened. The CLAHE parameters were heuristically tweaked to improve performance; a slighter larger tile size was found to improve results. One code repository[[1]](#footnote-1) was found to apply CLAHE only the luminance channel of a colour image to prevent unwanted hue and saturation change. Subsequent work could investigate the effect of this on calculating disparity.

**Disparity map calculation**

Two dense stereo approaches were compared. The first, Semi-Global Block Matching (SGBM), was introduced by Hirschmuller (2005). The second, Weighted Least Squares (WLS) uses two disparity maps – one for each camera – and combines them to produce a new map. The sub-processors used in WLS were also SGBM. The most space-expensive double-pass mode was used to improve performance in both filters. The image on the left is SGBM; on the right, WLS. WLS offers clearly superior performance: fewer blind spots and improved edge detection. However, “holes” in a disparity map can easily be ignored when computing distances.



Once the disparity map has been calculated, it is passed through a bilateral filter, introduced by Tomasi and Manduchi (1998), which smooths images while preserving edges. The image on the left is the original; on the right is the bilateral filtered version.



**Object detection**

Object detection used You-Only-Look-Once (YOLO), based on (Redmon and Farhadi, 2019). Sparse approaches were evaluated but deemed unsatisfactory. YOLO parameters were mostly left unmodified. The system frequently false detected trains, so these were manually filtered out. The system also frequently identified the car bonnet as a car, so this region was disabled. Modifying images prior to passing them into YOLO was found to reduce object detection accuracy. Non-maximum suppression worked well at 0.2. The confidence threshold seemed best around 0.55. Inference time is impressive: less than 500 ms on a fairly high-powered desktop. See the appendices for the annotated version of this image.

**Object distance calculation**

Several points of reference were established using Google Maps to estimate the expected distance between a detected object and the camera itself. These reference frames were evaluated after parameter modification. Bounding boxes entirely in the left-hand region of the image with no disparity information were discarded. The distance of boxes partially in the region was estimated using the available pixels. Various methods were tested for finding the distance. A Gaussian kernel was used to weight values at the centre of the bounding box more heavily than those at the outside, and this improve results, although occlusion remained a problem.

**Conclusion**

A reasonably accurate system has been developed. Results are consistent between different frames and of reasonable magnitude. A more sophisticated implementation would heuristically deal with occlusion or test the implemented stereo algorithm on a dataset with ground truth available, such as KITTI or Middlesbury, to better optimise parameters.

**References**

Ansar, A., Castano, A., & Matthies, L. (2004). Enhanced real-time stereo using bilateral filtering. *Proceedings. 2Nd International Symposium On 3D Data Processing, Visualization And Transmission, 2004. 3DPVT 2004.*. doi: 10.1109/tdpvt.2004.1335273

Deepa, & Jyothi, K. (2017). A robust and efficient pre processing techniques for stereo images. *2017 International Conference On Electrical, Electronics, Communication, Computer, And Optimization Techniques (ICEECCOT)*. doi: 10.1109/iceeccot.2017.8284645

Hirschmuller, H. (2005). Accurate and Efficient Stereo Processing by Semi-Global Matching and Mutual Information. *2005 IEEE Computer Society Conference On Computer Vision And Pattern Recognition (CVPR'05)*. doi: 10.1109/cvpr.2005.56

Kanade, T., Kano, H., Kimura, S., Yoshida, A., & Oda, K. (1995). Development of a video-rate stereo machine. *Proceedings 1995 IEEE/RSJ International Conference On Intelligent Robots And Systems. Human Robot Interaction And Cooperative Robots*. doi: 10.1109/iros.1995.525868

Redmon, J., & Farhadi, A. (2019). Yolov3: An incremental improvement. Retrieved from https://pjreddie.com/media/files/papers/YOLOv3.pdf

Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. *Sixth International Conference On Computer Vision (IEEE Cat. No.98CH36271)*. doi: 10.1109/iccv.1998.710815

Zuiderveld, K. (1994). *Graphics gems IV* (1st ed., pp. 474 -485). San Francisco, Calif.: Kaufmann.

**Appendices**

Source image



Median filter



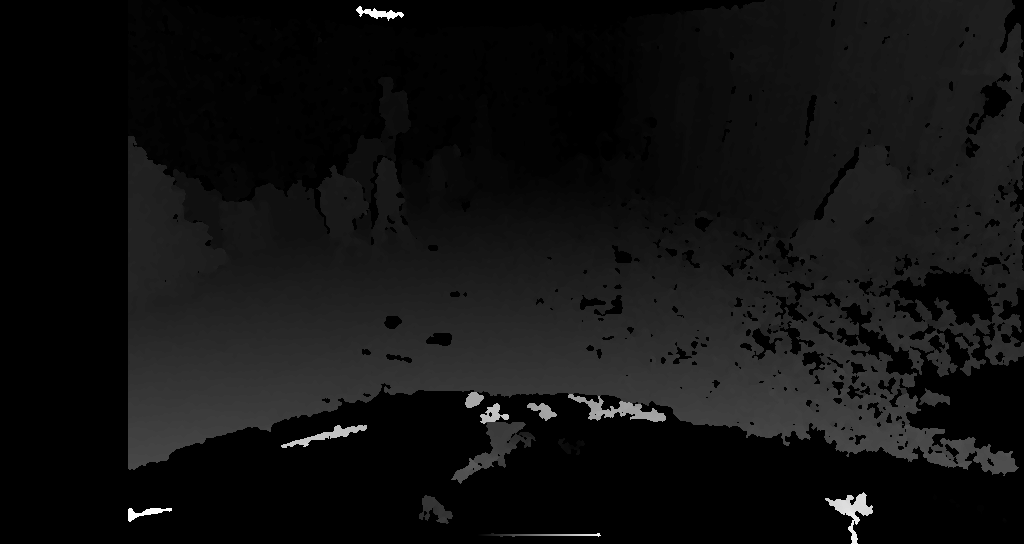
“Standard” histogram equalisation



CLAHE



SGBM



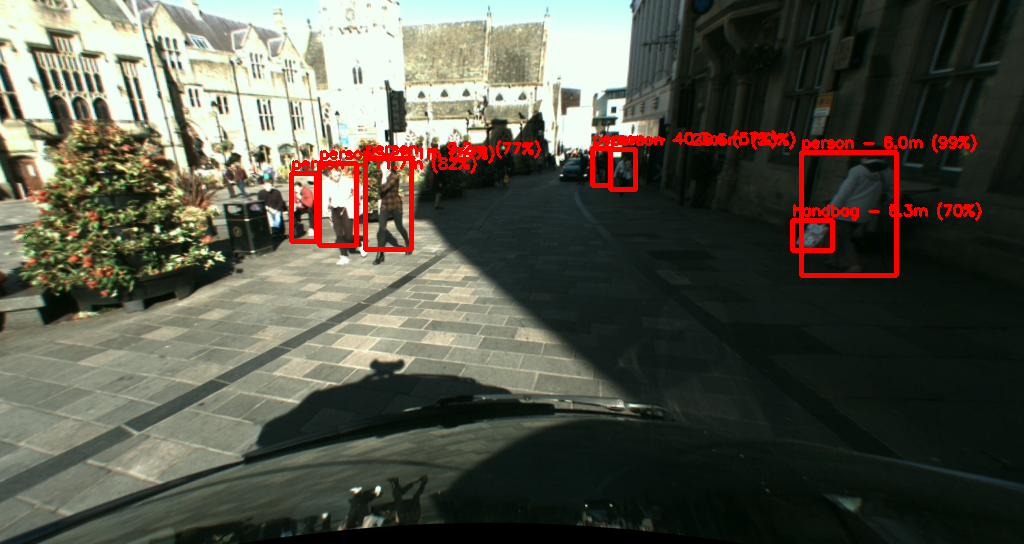
WLS



WLS, with bilateral filter



Original after YOLO



1. https://github.com/YuAo/Accelerated-CLAHE [↑](#footnote-ref-1)