Computer Vision Coursework

Stages: pre-processing, disparity map calculation, object detection, and object distance calculation.

**PRE-PROCESSING**

First step is to convert the images to greyscale.

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, and occlusion.

Scene changes in terrain type, illumination conditions, clutter, 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.



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 after 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.

**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.

INSERT FILTER COMPARISON HERE

**YOLO**

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.

Non-maximum suppression worked well. The confidence threshold seemed best around 0.55.

It is difficult to be objective about performance improvements.

“Holes” in a disparity map can easily be ignored when computing distances, and though subjective

Once the disparity map has been calculated, it is passed through a bilateral filter. The bilateral filter was introduced by Tomasi and Manduchi (1998). It smooths images while preserving edges. However, estimated distances rarely changed.

INSERT BILATERAL FILTER COMPARISON HERE

It is difficult to state accurate how quantitatively accurate the developed system is without a ground truth.

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.

The distance was estimated using the average of all the pixels. However, occlusion proved to be a problem.

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