1 Supplementary material

1.1 WIBAM dataset samples

(a)

(b)

Figure 1: Example images of the dataset showing the view from each camera. (a) Sample from the training set (b) Sample from the validation set

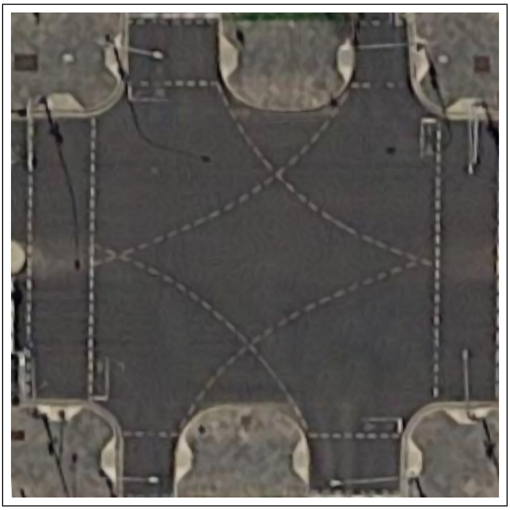


Figure 2: Satellite image of the intersection being observed. Imagery ©2021 Aerometrex Pty Ltd, Map data ©2021 Google

1.2 Camera synchronisation

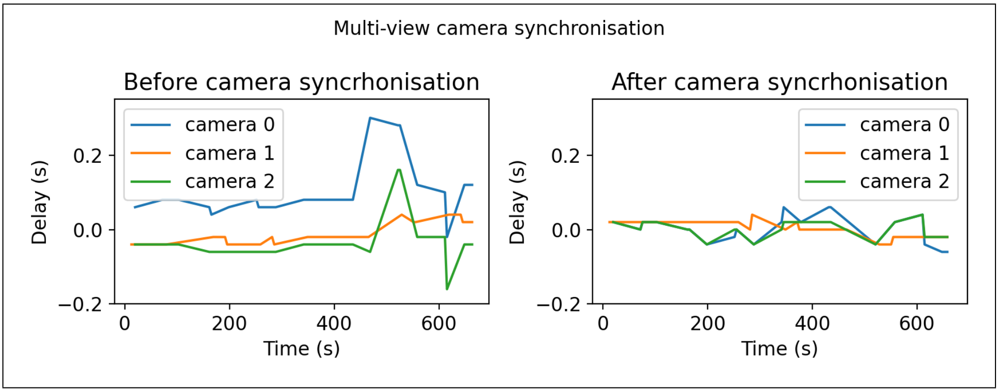
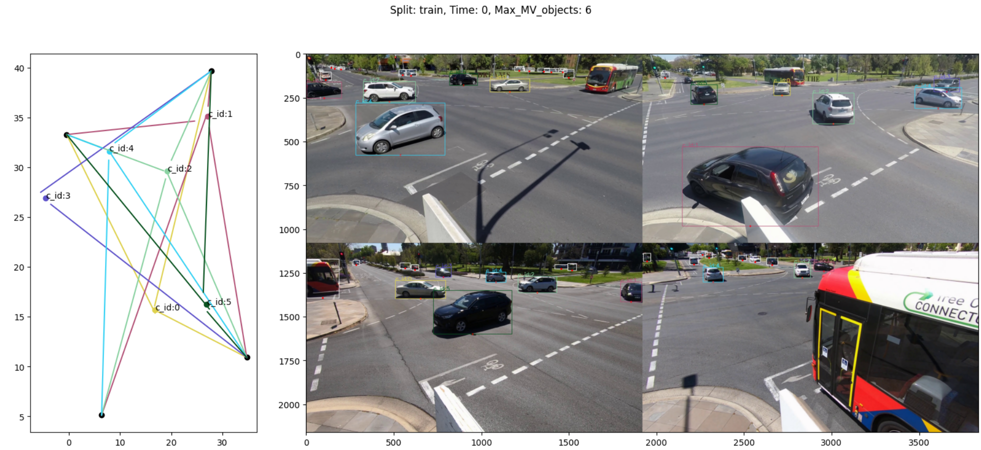


Figure 3: Before and after synchronising the dataset using the traffic lights changing. Delay with reference to camera 3 in the dataset.

Figure 3 shows the improvement in synchronisation between the multi-view cameras using our semi-automatic method. Before synchronisation of cameras, it is clear that there is a large error up to 16 frames towards the end of data collection. This large error can result in cars having zero overlap on the road surface between views. To measure the light changes we can use the change from green to yellow as the measurement signal and yellow to red light change as synchronisation signal. This measures the delay between cameras at the end of the light cycle and leaves a gap between red lights to measure if the cameras go out of sync again.

1.3 Clustering output



(a)

(b)

Figure 4: (a) Example of the clusters formed with vehicles detected from multiple views (b) Largest number (16) of clusters in the dataset. The colours correspond to a matched cluster. Black dots represent the four cameras with lines drawn out from the camera to the detection centre on the ground plane. Best viewed in colour and zoomed in.

Figure 4 shows the output of the clustering algorithm used for the multi-view vehicle associations. The diagram on the left shows the BEV clusters where the black dots are the camera centres with lines to the detection ground points. Objects which are clustered have a unique colour between the detection images on the right and the BEV diagram on the left. Figure 4(b) shows the highest number of matched vehicles the system found with 16 unique objects viewed at this time instance.

1.4 Automatically annotated multi-view vehicle distributions

Varying the positions of vehicles within the intersection give a wide range of depths and elevations vehicles are viewed from. There is also a different numbers of vehicles associated and matched across views at any given time, with a maximum of 16 shown in Figure 4(b). Figure 5 shows the distributions of these in the entire dataset. Our dataset views vehicles at a wide range of elevations, where autonomous vehicle data is typically around 5°.

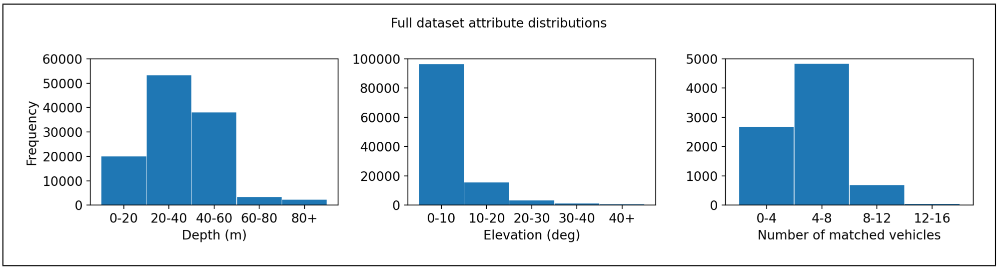


Figure 5: Distributions of key data elements within the WIBAM dataset

1.5 Test set distributions

Figure 6 shows the distributions of depth and elevation angle in the hand labelled test set of the WIBAM dataset. The distributions of elevations and distances are representative of the whole dataset however vehicles at larger distances outside of view for 4 cameras are difficult to label accurately hence there is a drop off above 40m+. The high mounted cameras result in less occlusions of vehicles therefore the majority are not occluded at all.

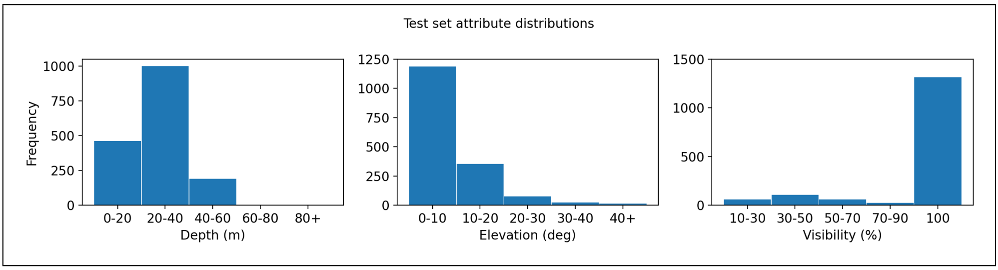
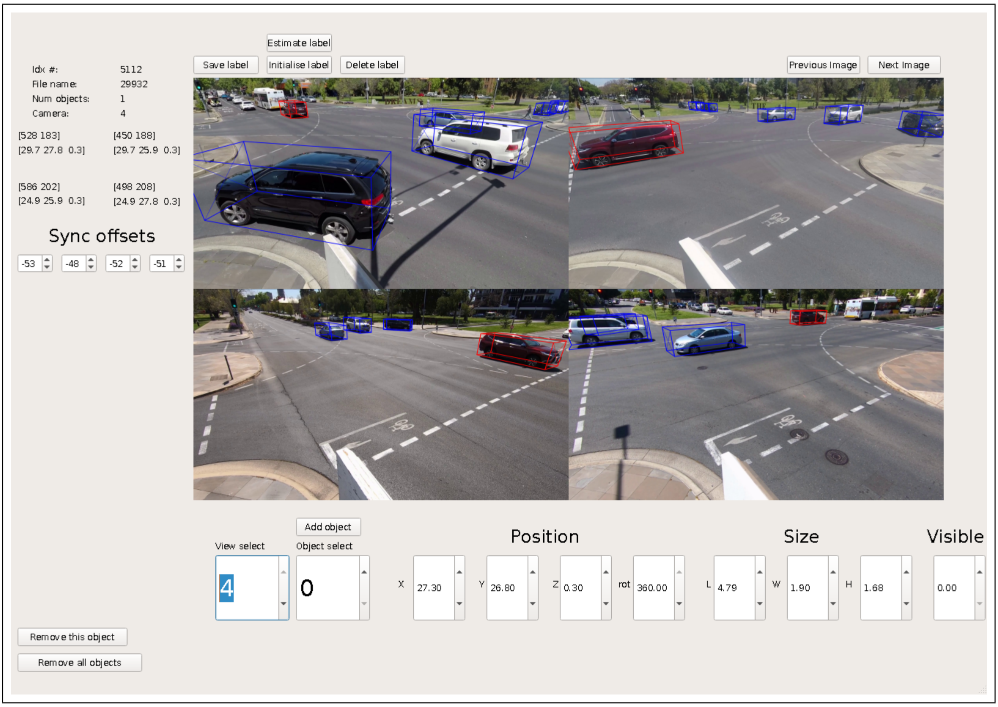


Figure 6: Hand labelled annotation distributions

1.6 Labelling tool GUI

The labelling tool allows the user to annotate vehicles with 7DoF pose. We found that annotating vehicles using this tool takes around 2.5 minutes per time instance or just over half a minute per image labelled.



(a)

Figure 7: The labelling tool created for labelling monocular multi-view data