# Generation of a Conversion Function for Distance Estimation

The objective is to generate a conversion function that takes as input the width in pixels from a detected vehicle and that provides as output the distance in meters to such detected vehicle. The following assumptions are made:

- The real-life width for all vehicles is approximately the same.

- The height of the camera (height above the ground plane, which is the road) will always be approximately the same.

- The camera will always have zero tilt (or constant tilt for all video captures if it is different from zero)

- The camera will always have zero pitch and zero roll angles. Considering the 3 rotation angles the ideal will be to have in 3D space the image plane of the camera parallel to the planes generated by the back-side of the vehicles in the road. Here it is also assumed that the surrounding vehicles move under usual conditions which are vehicles driving forward and parallel to the lane markings (zero degree heading).

- The road is a flat plane (almost universally valid) and the planes of the back-side of all the vehicles are orthogonal to the road's plane (obvious, nevertheless it needed to be said).

- Focal length provided by the manufacturer is a good representative average of the lens focal length.

- The CNN based object detection allows the generation of a bounding box that covers the entirety of the image content belonging to a vehicle. In this sense the bounding-box ideally covers the wheels of the vehicle, which provides a reference line that is very close to the 3D intersection between the vehicle's back-side plane and the road's plane. The intersection generates a 3D segment which works as a reference to determine distance.

- We assume that the vehicle has a real-world width of 1.8m. This assumption is valid for the greatest majority of vehicles in the road which are 4 wheel passenger vehicles. For buses and trucks this assumption can be changed accordingly.

The following equation provides the real world distance $D$ for a perceived width of the vehicle.

$$D = \frac{FW}{w}$$

Here $F$ is the camera's focal length, $W$ represents the real-world width of the vehicle, and $w$ represents the 2D perceived width of the vehicle. Using the equation above we can build a conversion function for different perceived widths of vehicles. In this case we can mount an ego-vehicle with a dashboard

camera and make marks on the pavement in front of the vehicle at specific longitudinal distances from the camera. The ego-vehicle is parked all the time. Then using another vehicle we can record a video of this vehicle as it moves slowly in front of the ego-vehicle and passes each one of the pavement markings. Then after processing the recorded video we measure the width in pixels of the vehicle at each of the distances indicated by the pavement marking where the vehicle is located. This will provide us with a function that maps vehicle widths with real world distances. Such function can be used to directly obtain longitudinal distance estimation on deployment. This function can be also used as an independent method of corroboration of the accuracy of the longitudinal distance estimation method we presented in the corresponding paper "Real-time Automotive Risk Estimation Using Monocular Video".