

Automated Camera Stabilization and Calibration for Intelligent Transportation Systems

Marcel Bruckner
 Technische Universität München
 Boltzmannstraße 15, 85748 Garching
 marcel.bruckner@tum.de

0.1. Static calibration

Intelligent Transportation System are inherently dependent on the calibration of the different sensors. To track and predict traffic the system has to know the poses of the different sensors relative to some reference coordinate system. This enables the ITS to accurately measure the position of vehicles within the single sensor ranges and at the overlapping boundaries.

Previous experiments have shown that a calibration process based on an IMU is not feasible in our case. Instead we focus on a calibration procedure based on visual landmarks in the video feed. The landmarks are mapped to their partially known world positions from high definition road maps.

Retrieve objects positions from the HD maps (??) In this work we focus on the permanent delineator objects that are easily visible in the video feeds.

The world position of the objects can be retrieved using the mathematical operations defined in the OpenDRIVE standard.

This gives us the the base origin point $o = (x, y, z)^T$ of the object in the transverse mercator projection [1]. The base origin point is the world position of the lower end of the object where it ends in the ground or another object.

Additionally we retrieve a directional heading axis $d = (x, y, z)^T$ and the height h of the object.

These values enable us to approximate the real-world objects by sampling points $s \in S$ in world position along the center line of the object, where

$$s = \{o, d, \lambda\} \in S, \quad \lambda \in \Lambda, \quad \Lambda = \mathbb{R} \quad (1)$$

and defines a world position calculated by

$$w(s) = s_o + s_\lambda * s_d \quad (2)$$

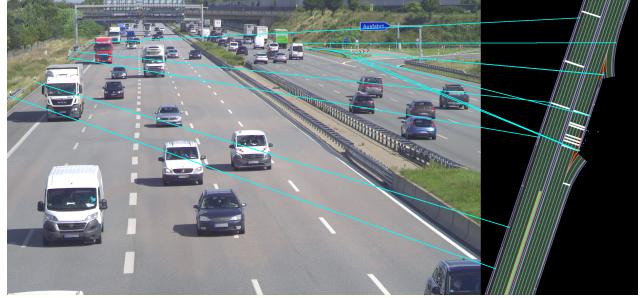


Figure 1

Mapping pixels to objects To calibrate the camera we need to establish a mapping

$$p_c \sim s_c \quad (3)$$

from pixels $p_c = (u, v)^T \in P$ from the video stream to the corresponding sampled points s_c from the object they belong to.

Calibration procedure Our camera is modelled using the pinhole camera model. This uses the known intrinsic camera parameters and their corresponding projection π from camera to image space.

The extrinsic camera parameters that map from world space to camera space is defined by the camera translation T and the camera rotation R . The translation and rotation are unknown, for them we optimize.

We estimate the camera pose by minimizing a modified version of the reprojection error

$$\min_{T, R, \Lambda} E(P, S, T, R, \Lambda) \quad (4)$$

formulated as

$$E(P, S, T, R, \Lambda) = \sum_c \left\| \begin{pmatrix} p_c - \pi(T * R * w(s_c)) \\ clamp(\lambda, 0, h) \end{pmatrix} \right\|_2^2 \quad (5)$$

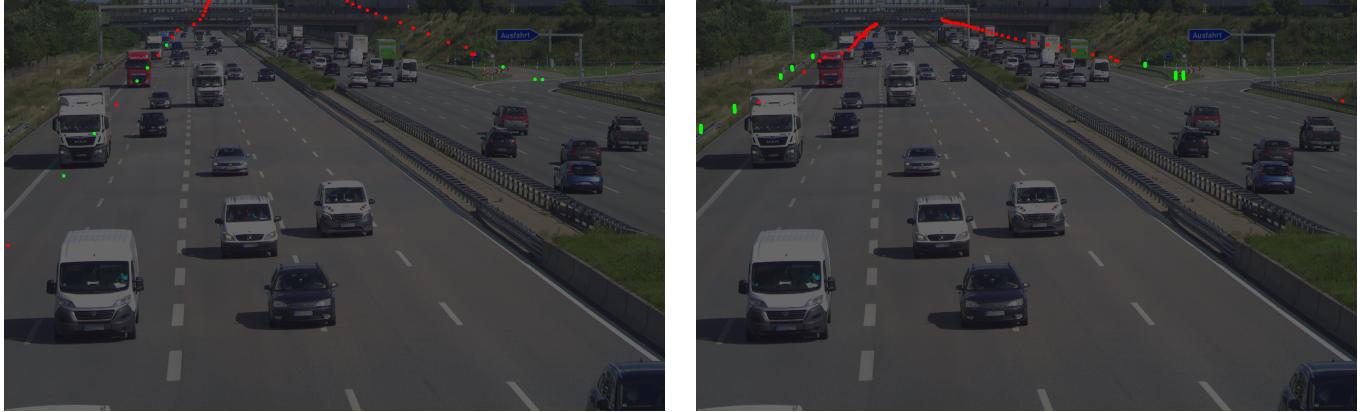


Figure 2: This is the calibration

where P is the set of mapped pixels in the image, S is the set of mapped corresponding sampled points from the objects and Λ is the set of λ values associated with the sampled points and

$$clamp(x, a, b) = \begin{cases} x - b, & \text{if } x > b \\ x, & \text{if } x < 0 \\ 0, & \text{else} \end{cases} \quad (6)$$

The world positions of the objects lie on the one-dimensional lines that approximate the objects we have to also optimize for the actual λ values of the samples. This jointly finds the world position of the samples and the pose of the camera.

References

- [1] PROJ contributors. *PROJ coordinate transformation software library*. Open Source Geospatial Foundation, 2021. [1](#)