Automated Camera Stabilization and Calibration for Intelligent Transportation Systems

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

1. Terms and Definitions

This section provides explanations of extensively used terms and definitions.

1.1. Digital Twin

1.2. Dynamic foreground and static scene

We group parts of the world seen by the cameras into the two groups of dynamic foreground and static scene. Dynamic foreground describes all pixels that represent objects that are meant to be moving, *e.g*. vehicles on the roads. On the other hand the static scene are all pixels that are not dynamic foreground, *e.g*. the road, guardrails or bridges.

1.3. Parametric cubic polynomial

To calculate a parametric cubic polynomial the following equation is used.

$$\begin{aligned} para(U,s) &= para(U_a,U_b,U_c,U_d,s) \\ &= U_a + U_b * s + U_c * s^2 + U_d * s^3 \end{aligned} \tag{1}$$

1.4. 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.

High definition road maps (HD maps) Several HD maps are used to generate the Digital Twin (1.1). The HD maps adhere to the OpenDRIVE standard V1.4. They contain spatial and relational information between the world, roads, the road coordinate frames and objects within these frames.

Retrive 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 $h = (x,y,z)^T$ and the height λ of the object.

These three values enable us to approximate the real-world objects by sampling points s in world position

$$s \in \{s | s = p + \mu * h : \mu \in [0, \lambda]\} \tag{2}$$

along the center line of the object.

Mapping pixels to objects

References

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